### Title

Sibling Constructs: What Are They, Why Do They Matter, and How Should You Handle Them?




### Journal

Personality and Social Psychology Review, 25(4)

### ISSN

1088-8683



### Authors

Lawson, Katherine M

Robins, Richard W



### Publication Date

2021-11-01

### DOI

10.1177/10888683211047101

Peer reviewed

*Article*

# Sibling Constructs: What Are They, Why Do They Matter, and How Should You Handle Them?

Personality and Social Psychology Review
2021, Vol. 25(4) 344–366

DOI: 10.1177/10888683211047101
pspr.sagepub.com
SAGE

**Katherine M. Lawson[1] and Richard W. Robins[1]**

**Abstract**

Researchers often study constructs that are conceptually and/or empirically related, but distinct (i.e., "sibling constructs"). In social-personality psychology, as well as psychology more generally, there is little guidance for how to deal with sibling constructs, which can result in researchers ignoring or mishandling them. In this article, we start by situating sibling constructs in the literature on the jingle-jangle fallacies. Then, we outline 10 conceptual and empirical criteria for determining the degree to which, and in what ways, constructs may share a sibling relationship, using self-esteem and grandiose narcissism as a running example. Finally, we discuss strategies for handling sibling constructs in a systematic and transparent way. We hope that the procedures described here will help social-personality psychologists identify sibling constructs, understand when and why they pose problems for their research, and adopt strategies that ameliorate their adverse effects.

**Keywords**

sibling constructs, jingle-jangle, theory, self-esteem, narcissism

The jingle-jangle problem permeates psychological research. First discussed over a century ago, the "jingle" fallacy occurs when researchers assume that two distinct constructs are the same because they bear the same name, resulting in an "unthinking acceptance of verbal equality as proof of real equality" (Thorndike, 1904, p. 14). The "jangle" fallacy occurs when researchers assume that two identical, or nearly identical, constructs are different because they have different labels (Kelley, 1927). The jingle-jangle fallacies can impede scientific progress; as Block (2000) noted, "the jingle fallacy can cause the scientifically uncareful to believe they are talking of the same phenomenon when indeed they are not," whereas the jangle fallacy "limits discernment of important empirical and conceptual convergences" (p. 156). In short, the jingle-jangle fallacies "waste scientific time" (Block, 1995, p. 2010).

Further complicating matters, researchers sometimes discuss the jingle-jangle fallacies with respect to the name of a *construct* (e.g., are grit and conscientiousness really the same construct?), and other times with respect to the name of a *measure* (e.g., does the Grit Scale [Duckworth et al., 2007] measure the same thing as the Conscientiousness scale from the Big Five Inventory measure of conscientiousness [Soto & John, 2017]?). Indeed, these fallacies exist at both the construct *and* the measure level, an important distinction that is rarely made explicit. As we will discuss below, jingle-jangle problems at the construct level tend to create problems at the measurement level, and vice versa, leading researchers to get even more lost in the "jingle-jangle jungle" (Block, 1995; Peck, 2004).

Given the pervasive and perplexing nature of the jingle-jangle problems, they have received a great deal of attention. A Google Scholar search revealed more than 3,300 psychology articles with jingle-jangle in the title. In these articles, researchers often describe the conceptual gray area that jingle-jangle creates, using words like "conceptual haziness" (Reschly & Christenson, 2012), "conceptual fuzziness" (Weidman et al., 2017), and "murky distinction" (Marsh et al., 2019). These descriptions allude to the prevalence of muddled distinctions between constructs (or measures)—what Block (1995) referred to as the "hasty, hazy, lazy use of language" (p. 209)—which further exacerbate jingle-jangle confusion. Indeed, jingle occurs when two constructs or measures are distinct, but named the same, whereas jangle occurs when two constructs or measures are identical, but named differently. But what makes two constructs *different enough* to be considered distinct? What makes two constructs *similar enough* to be considered identical? What happens with constructs that fall into the gray area between distinct and identical? And, most important, what steps should researchers take to ensure that constructs coexisting in this gray area do not lead to theoretical and methodological confusion?

[1]University of California, Davis, USA

**Corresponding Author:**
Katherine M. Lawson, Department of Psychology, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA.
Email: kmlawson@ucdavis.edu

This article attempts to address these thorny questions. Specifically, we examine a pervasive phenomenon in psychology where researchers are dealing with constructs that are conceptually and/or empirically related, but distinct. We will call these overlapping but distinct constructs "sibling constructs." This term captures the fact that the constructs share a close, familial relation, but are not identical; that is, they are not "twin constructs." Of course, in the field of psychology, we generally assume that all cognitive, affective, interpersonal, and behavioral processes are part of the same interconnected system, and therefore, all psychological constructs are interrelated to some degree. This is one reason why Meehl (1967) claimed that the null hypothesis specifying no relation between psychological variables is nearly "always false" (p. 108). However, even if we accept Meehl's dictum, researchers must still determine how distally or proximally connected the constructs are to elucidate whether they are siblings or merely distant cousins.

Given the prevalence of sibling constructs in psychological research, one might expect that researchers would have a systematic method for identifying and handling them. However, there is surprisingly little guidance, with norms varying widely across research areas. Consequently, researchers often fail to identify sibling constructs or ignore any problems that might arise from their existence. When researchers do recognize and attempt to address the problems associated with a strong sibling relationship (e.g., by controlling for shared variance between measures), they often do so in varied and capricious ways, with little justification for the approach taken. Thus, the field needs more systematic and standardized strategies for handling them.

We begin by discussing the importance of identifying sibling constructs and the problems they often create for an area of research. Next, we outline 10 criteria for determining the degree to which, and in what ways, constructs are siblings, and describe conceptual and statistical methods to evaluate each criterion. To illustrate how these criteria can be applied in an actual research context, we discuss numerous examples of sibling constructs from different areas of social-personality psychology and include a running example of self-esteem and narcissism. In the final section, we provide several broad recommendations for how researchers should deal with sibling constructs to reduce their potential negative impact on research.

## Why Should You Care About Sibling Constructs?

The field of social-personality psychology is replete with constructs that have some degree of conceptual and/or empirical overlap, including self-esteem and narcissism, implicit and explicit prejudice, racism and sexism, shame and guilt, patriotism and nationalism, and the set of traits and processes subsumed under the umbrella term "self-regulation." Some of these sets of constructs have "content domains captured by different measures [that] are so strongly overlapping that treating them as being separate become[s] fairly dubious" (Leising et al., in press, p. 12). Others are conceptualized similarly, but empirical measures between them barely correlate. These constructs pose numerous problems for researchers, in part because they exist in a crowded but often ignored "gray area" of the jingle-jangle jungle. Peering through the thick trees, how is one to know whether the overlap between two constructs has crossed the threshold to being "fairly dubious"? Jingle-jangle problems are most often discussed in categorical terms—that is, two constructs have either been jingled/jangled or they have not. But many relations between constructs do not fit neatly into those boxes, which leads some researchers to treat them as entirely distinct, others to recognize their conceptual and/or empirical overlap and take steps to address it, and others to lump them together as one and the same construct (and consequently, assume that findings for one necessarily hold for the other). Far too often, these divergent ways of conceptualizing and handling sibling constructs are based on varied and nonsystematic forms of evidence, despite the fact that each approach—treating them as entirely different, overlapping, or the same—has profound consequences for how they are studied and how research on the constructs is interpreted (or misinterpreted).

Overlooking jangled constructs may lead researchers to propose new constructs and/or measures that already exist in the literature, thus wasting time, journal space, and research funding to further muddy the waters. Some researchers argue that this happened with grit and conscientiousness. Grit is defined as "perseverance and passion for long-term goals" (Duckworth et al., 2007, p. 1087) and involves putting in sustained effort toward goals even in the face of adversity. Conscientiousness, one of the Big Five personality domains, involves "the propensity to be self-controlled, responsible to others, hardworking, orderly, and rule abiding" (Roberts et al., 2014, p. 1). Self-report measures of grit and conscientiousness are highly correlated (e.g., $r = .77$; Duckworth et al., 2007) and their theoretical conceptualizations are quite similar. Duckworth and colleagues (2007) argue that, unlike conscientiousness, grit emphasizes long-term stamina and consistent goals and interests, but others note that (a) the conscientiousness domain includes persistence toward long-term goals and other aspects of grit, and (b) measures of the two constructs seem to assess the same underlying trait (Credé et al., 2017; Ion et al., 2017). This example demonstrates an instance when both constructs *and* measures have been jangled, resulting in separate literatures that should be unified.

In contrast, overlooking jingled constructs may lead researchers to mistakenly conclude that a finding is not replicable or that there are boundary conditions, when actually the inconsistent results reflect important differences between the constructs (Lilienfeld & Strother, 2020). For example, executive function was long thought to be a broad, but

coherent, construct (Miyake et al., 2000). This assumption led to a morass of inconsistent findings because measures of executive function tap into distinct facets of what we now know is not a single, unitary construct, but rather a collection of loosely related, but dissociable, processes, including working memory, inhibitory control, and cognitive flexibility (Burgess et al., 1998; Kramer et al., 1994; Miyake et al., 2000). This example highlights how failing to notice jingled constructs can lead to years of nonreplicable research (e.g., Blair et al., 2005), and widespread confusion about the nature of the construct(s).

Similarly, Baumeister and colleagues (1996) failed to distinguish (i.e., jingled) self-esteem and narcissism (Donnellan et al., 2005). Self-esteem refers to an individual's "subjective evaluation of his or her worth as a person" (Donnellan et al., 2013, p. 718), whereas narcissism involves a tendency to be "self-promotional, self-enhancing, and entitled" as well as antagonistic and extraverted (Miller et al., 2017). However, Baumeister and colleagues (1996) treated these constructs as interchangeable, leading them to argue (based largely on research examining narcissism) that high self-esteem is associated with violence and aggression, and conclude that promoting high self-esteem might be "counterproductive and even dangerous" (p. 29). Together, these examples highlight how misunderstandings and misinterpretations about sibling constructs can waste researcher time, impede the accumulation of scientific findings, and contribute to widespread issues like the replicability crisis.

Instead of trying (and sometimes failing) to make black-or-white classifications for constructs and measures that fall in the gray area of similarity versus distinctness (i.e., are they the same or different?), we propose that researchers should instead embrace the continuum. That is, recognize that constructs and measures exist on a spectrum of similarity—ranging from extremely similar to moderately similar to only slightly similar—and pursue strategies aimed at understanding the *degree* of similarity and documenting the *ways* in which they are similar and distinct.

This approach will alleviate many of the problematic consequences described above. Instead of one research group advocating that grit and conscientiousness are distinct and another group trying to prove that they are identical, both groups can focus their efforts on systematically documenting the specific ways in which the constructs and measures overlap and diverge. Similarly, instead of selectively choosing findings that suggest similarities between self-esteem and narcissism and then representing them as identical, researchers would have to systematically consider a wide range of conceptual and empirical criteria—some of which might provide evidence for their similarities and others that might highlight their differences. Embracing our approach will also enhance collaboration across research groups because, like formalized theories, assumptions will be clearly explicated and shared transparently (e.g., see Millner et al., 2020).

## Identifying Sibling Constructs

The first step toward dealing with sibling constructs is establishing systematic procedures for identifying them and exploring their similarities and differences. In most areas of research, decisions about whether constructs should be treated as siblings occur through a haphazard process, with different researchers drawing on different forms of evidence to bolster their position. Thus, the field would benefit from a set of structured criteria that will reduce unfounded claims and allow researchers who disagree about the sibling status of constructs to pinpoint areas of contention.

Below we propose 10 criteria that we believe encompass the most important ways in which constructs and measures can be related to each other. Specifically, a sibling relationship between two (or more) constructs can be established considering the extent that they (1) are defined in a conceptually similar way; (2) have a high degree of overlap in their theorized nomological networks; (3) have a high degree of overlap in their observed nomological networks; (4) have measures that correlate strongly with each other; (5) have measures that together form a strong general factor; (6) have measures that show little incremental validity over each other; (7) have similar developmental trajectories; (8) share underlying causes (including environmental causes, genetic variance, and neural mechanisms); (9) are causally related to each other; and/or (10) are state/trait manifestations of the same underlying process (see Table 1). Notably, application of these criteria will not produce a clear categorical judgment about whether or not two constructs are the same or different, but rather they provide a framework for determining the degree and nature of the overlap between constructs. This approach to understanding sibling relationships is similar to the process of validating a scale, in that it involves a never-ending accumulation of evidence about the meaning of each construct vis-à-vis the other. We believe that the widespread application of these criteria will help reduce the density of the jingle-jangle jungle, raise the bar for introducing new constructs and measures in the field, and lead to improvements in research on sibling constructs. The first two criteria—conceptual similarity and overlapping theorized nomological networks—refer to theoretical considerations about the *construct*. The next four criteria focus on the *measure* of a construct and therefore require both theoretical and empirical considerations. The last four criteria concern how constructs *emerge, develop, and change over time*, taking advantage of methods that capture longitudinal associations to identify siblings.

### *1. Conceptual Similarity*

Constructs are conceptually related when they involve a similar underlying trait or process. When researchers have clear definitions of each construct, they can identify substantial

**Table 1.** Ten Criteria for Identifying Sibling Constructs and Exploring Their Similarities and Differences.

| Criterion | Category | Method for Evaluating Criterion |
|---|---|---|
| 1. Defined in a conceptually similar way | Conceptual | Identify/formulate explicit conceptual definition of each construct, then compare |
| 2. High degree of overlap in their theorized nomological networks | Conceptual | Create nomological network for each construct, then compare |
| 3. High degree of overlap in their actual nomological network | Empirical | (a) Examine convergent/divergent correlations between measures of both siblings and a set of outcomes that vary in relevance to the constructs; (b) quantify similarity in nomological networks by examining profile agreement (intraclass correlations) between the convergent/divergent correlations found for each construct |
| 4. Strong association between measures of siblings | Empirical | (a) Correlate measure of one sibling construct with a measure of the other sibling construct; (b) estimate a multitrait-multimethod matrix with multiple measures of each sibling construct and examine strength of trait vs. method factors |
| 5. Sibling measures form a strong general factor | Empirical | (a) Conduct a confirmatory factor analysis to test whether a 2-factor model (representing the two sibling constructs) fits significantly better than a single factor model; (b) estimate a hierarchical or bifactor model and examine the variance accounted for by the general versus lower-order specific factors as well as model fit before and after the lower-order factors are specified. |
| 6. Sibling measures show little incremental validity over each other | Empirical | Enter measures of each sibling in a multiple regression and see whether they have incremental validity in predicting relevant outcomes. |
| 7. Similar developmental trajectories | Developmental | Examine whether constructs show evidence of codevelopment by examining strength of association between developmental trajectories using longitudinal data, for example, using bivariate latent growth curve models or FOCUS models |
| 8. Shared underlying causes | Developmental | Examine empirical evidence for shared environmental antecedents, underlying neural mechanisms, and genetic variance |
| 9. One sibling causes the other | Developmental | (a) Use experimental manipulation and/or interventions to test if one variable causes the other, or whether you can independently manipulate one sibling construct without also changing the other sibling construct; (b) use cross-lagged panel models (both traditional and random intercept) to examine whether there are significant lagged effects between siblings (but understand the challenges of inferring causation from correlational data; Rohrer, 2018); (c) use DAGs to account for the presence of confounders (Pearl, 2009) |
| 10. State/trait manifestations of the same underlying process | Developmental | (a) Examine whether conceptualizations of constructs are similar processes just on different time scales; (b) use measures on different time scales (e.g., ask about state vs. trait manifestations of construct) and examine whether and how scores on both converge/diverge |

DAGs = directed acyclic graphs; FOCUS = factor of curves.

conceptual similarity by assessing similarities between these definitions. Problems arise, however, when this seemingly simple criterion is applied to actual constructs in psychology because definitions of constructs are often vague and incomplete and there may be multiple competing conceptualizations of the same construct. Therefore, to apply this criterion, there must first be a specific, comprehensive definition of each construct. Beyond defining each construct, this criterion also requires conceptualizing how the constructs relate to each other. For example, are the two constructs related because both are facets of the same superordinate domain (i.e., a parent factor), or because one is a facet of the other? Is one simply a more extreme (or more maladaptive) version of the other, but they are arrayed on the same continuum? Are they two

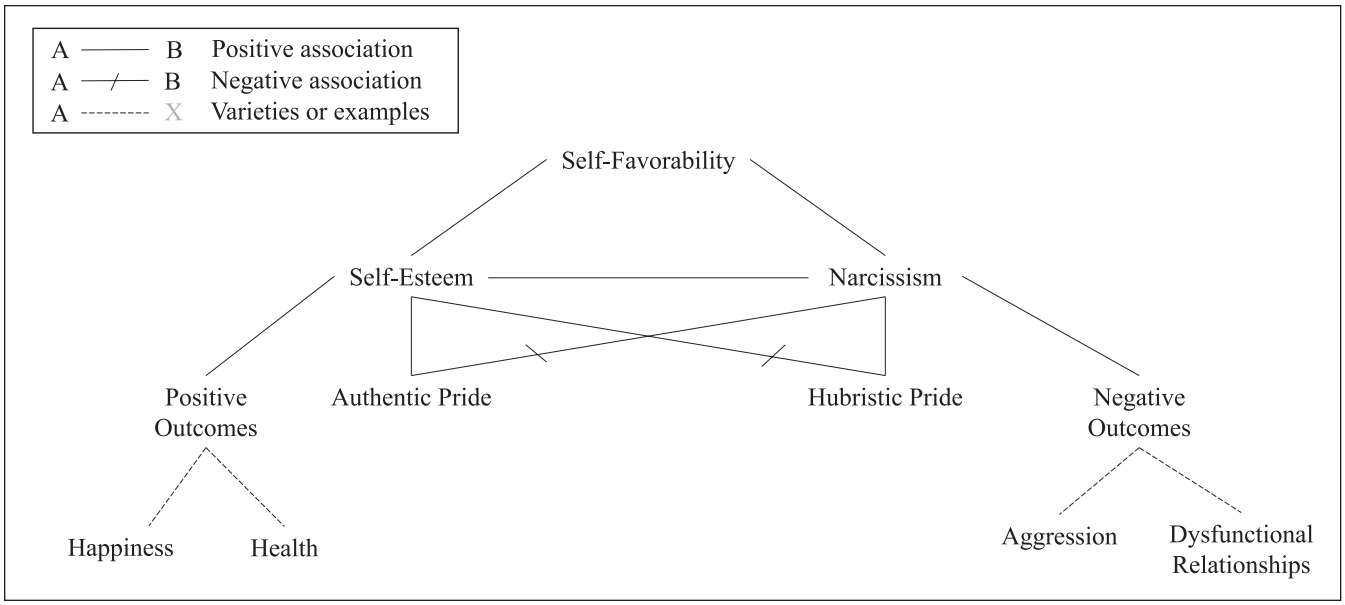


**Figure 1.** Example theory map of self-esteem and narcissism.
*Note.* This theory map uses guidelines outlined by Gray (2017). It is a simplified version of a comprehensive theory map for self-esteem and narcissism.

different manifestations of essentially the same underlying construct, such as implicit versus explicit, or state versus trait, expressions of a construct? Clarifying these conceptual issues is critical when dealing with sibling constructs.

Take implicit and explicit prejudice as an example. Conceptually, explicit prejudice reflects a person's conscious attitudes about another group, whereas implicit prejudice reflects unconscious attitudes that operate largely outside of the person's awareness (Dovidio et al., 2002). Despite operating at different levels of consciousness, both constructs share conceptual similarity because they both reflect prejudicial attitudes toward another group. Similarly, racism and sexism are conceptually similar because they both involve a general prejudicial attitude toward others, but with the target of prejudice varying (i.e., racial/ethnic minorities vs. women; Swim et al., 1995).

Self-esteem and narcissism are both characterized by positive evaluations of the self; that is, both self-esteem and narcissism come from the "parent" trait of self-favorability (Paulhus et al., 2004).[1] However, there are important conceptual differences between them (Brummelman et al., 2016; Hyatt et al., 2018; Paulhus et al., 2004; Tracy et al., 2011). Specifically, self-esteem is often conceptualized as the extent to which individuals like themselves and see themselves as worthy regardless of actual abilities or social acceptance, whereas narcissism is often conceptualized as the extent to which a person is "arrogant, egotistical, or has an otherwise grandiose sense of self" (Donnellan et al., 2013, p. 720). This distinction is consistent with other theoretical conceptions of the constructs, including Rosenberg's (1989) view that high self-esteem (i.e., feeling positively about one's competence adequacy and worth) is not synonymous with egotism (i.e., feeling superior to others). Thus, self-esteem and narcissism share some conceptual similarity, but not enough (in our view) to be considered twin constructs (i.e., they have not been jangled).

## *2. High Degree of Overlap Between Theorized Nomological Networks*

Constructs can also overlap conceptually, and therefore share a sibling relationship, because they share similar theorized nomological networks. A nomological network is a representation of interrelations between a construct of interest and various other constructs, variables, and observable properties (Cronbach & Meehl, 1955). Creating a theorized nomological network requires careful conceptual consideration, given the wide range of psychological attributes and processes that might be associated with a particular construct. One helpful method for outlining a nomological network is to create a theory map, which involves visually depicting associations among constructs, including moderators of associations and specific varieties and examples of the constructs (Gray, 2017; see Figure 1 for a simplified example of a theory map). Sibling constructs can vary in how much they overlap (i.e., converge and diverge) in their theorized nomological networks. The extent of the overlap will depend on various characteristics of the sibling constructs including the amount that is known about a construct (Cronbach & Meehl, 1955) and the breadth of the construct's nomological network (Lynam et al., 2006). Broader constructs with far-reaching nomological networks (e.g., the Big Five domains) are more

likely to overlap with other constructs, whereas more narrow constructs with less extensive nomological networks are less likely to overlap with other constructs.

Consider executive functions as an example. Despite sharing little conceptual similarity (e.g., working memory captures immediate conscious perceptual processing, whereas inhibitory control involves overriding a dominant response to instead choose a more appropriate behavior), the various executive functions have broad nomological networks that are theorized to overlap with each other. For example, executive functions are thought to share similar associations with fluid intelligence (Arffa, 2007) and attention deficit hyperactivity disorder (ADHD) symptoms (Nigg et al., 2002). This overlap between the theorized nomological networks of various executive functions indicates the presence of a broader parent construct and speaks to their standing as siblings.

Self-esteem and narcissism are conceptually similar, yet there are clear distinctions between their theorized nomological networks (see Figure 1). High self-esteem is theorized to be associated with positive life outcomes like health and happiness, whereas narcissism is thought to be associated with negative outcomes like aggression and failed interpersonal relationships (W. K. Campbell et al., 2002; Falkenbach et al., 2013; Locke, 2009; Orth & Robins, in press; Tracy et al., 2009). Furthermore, self-esteem is thought to be associated with feelings of authentic pride, which is marked by a sense of accomplishment and mastery (Tracy & Robins, 2007), whereas narcissism is theoretically associated with arrogant and conceited hubristic pride (Tracy et al., 2009). Moreover, Brummelman et al. (2016) suggest that individuals high in narcissism care more about getting ahead, whereas those high in self-esteem care more about getting along. Thus, despite sharing some conceptual overlap, self-esteem and narcissism have important, nonoverlapping aspects of their theorized nomological networks.

The two criteria discussed so far—conceptual similarity and overlap of theorized nomological networks—are purely theoretical. The next set of criteria require empirical methods to identify sibling constructs. These criteria are more complex to assess because researchers must be mindful of considerations at both the construct and measure levels. That is, researchers must have clear conceptualizations of the constructs they are studying and a clear understanding of the psychometric properties of the measures they are using to operationalize these constructs. Therefore, the following criteria require the use of measures with adequate construct validity (Cronbach & Meehl, 1955; Flake et al., 2017; Loevinger, 1957); without construct valid measures, these criteria cannot be meaningfully applied. This is especially important with sibling constructs, where the corresponding measures need to validly assess each construct *and* validly discriminate between the two constructs. Thus, for all the remaining criteria, the results of the proposed analyses of sibling relationships should be interpreted in the context of the psychometric properties of the measures. This is not an easy recommendation to follow because debates about whether two constructs are siblings often reflect debates about the validity of the measures used to study them. As noted earlier, research on the degree to which and the ways in which constructs are siblings is analogous to the process of construct validation, except that the validation process is happening in tandem for the measures.

### *3. High Degree of Overlap Between Empirically Documented Nomological Networks*

For the previous criterion, we considered overlap between the *theorized* nomological networks of the two potential sibling constructs. For this criterion, we consider the overlap between the actual (i.e., *observed)* pattern of convergent and discriminant associations in the nomological network. Researchers can accomplish this by examining the degree to which the potential siblings have a similar pattern of observed correlations with more (vs. less) theoretically relevant constructs.

Consider the literature on shame and guilt as an example. Shame and guilt are both self-conscious emotions elicited by negative events, but shame is elicited by negative attributions about the self (e.g., "I am a bad person"), whereas guilt is elicited by negative attributions about one's behavior (e.g., "I did a bad thing"; Tracy & Robins, 2006). Tangey and colleagues (1996) argue that shame and guilt, despite arising in similar situations, are "distinct emotions with different phenomenological features, different ways of experiencing interpersonal contexts, different ways of construing the emotion-eliciting behavior or situation, and different motivations for subsequent action" (p. 1264). To facilitate research on shame and guilt, Tangney et al. (1989) developed the TOSCA, which was specifically designed to tease apart the theoretical distinctions between the two emotions, by forcing participants to choose either a shame *or* a guilt response (but not both) when presented with various emotion-eliciting scenarios. Research using the TOSCA has demonstrated that shame and guilt are associated with similar appraisal processes (Tracy & Robins, 2006), but shame is negatively associated with adaptive outcomes like well-being and self-esteem, whereas guilt is positively associated with these adaptive outcomes (Tangey & Dearing, 2002). Thus, their observed nomological networks appear to be quite distinct.

Similarities and differences in convergent and divergent associations of self-esteem and narcissism further support their status as sibling constructs. In particular, Hyatt et al. (2018) found that both self-esteem and grandiose narcissism had positive convergent correlations with agentic traits (e.g., extraversion, skill), assertive interpersonal approaches, and positive affect, demonstrating areas of overlap. However, these authors also found substantial differences in associations, such that self-esteem was significantly associated with good parenting experiences (e.g.,

warmth, monitoring), fewer reports of abuse, and lower negative affect, whereas narcissism had small or nonsignificant associations with these variables. There are also documented differences between the emotional responses of individuals with high levels of self-esteem (associated with the experience of authentic pride) and those with high levels of narcissism (associated primarily with hubristic pride; Dickens & Robins, 2020; Tracy et al., 2009, 2011). Individuals with high levels of genuine self-esteem (i.e., self-esteem controlling for narcissism) show fewer aggressive and antisocial behaviors, whereas individuals prone to narcissistic self-aggrandizement (i.e., narcissism controlling for self-esteem) report higher levels of aggression and antisocial behavior (Tracy et al., 2009). Moreover, self-esteem is more closely tied to feelings of guilt, but not shame, whereas narcissism shows the opposite associations (Tangey & Dearing, 2002; Tracy et al., 2009). Overall, self-esteem tends to be positively correlated with well-being and adjustment, whereas narcissism tends to be weakly or uncorrelated with measures of well-being and adjustment (Baumeister et al., 2003; Tracy et al., 2009). Thus, self-esteem and narcissism share both similarities and differences in their patterns of convergent and divergent associations with numerous other variables.

So far in this section, we have been making qualitative judgments about similarity between nomological networks. However, there are quantitative methods for comparing nomological networks. Measures of profile agreement like the double-entry intraclass correlation ($ICC_{DE}$; McCrae, 2008) can quantify the extent to which two siblings share similar "profiles" of convergent and divergent correlations with the variables in their shared nomological networks. Higher $ICC_{DE}$ values (e.g., $ICC_{DE} = .60$) would indicate the siblings share similar associations with variables in their nomological network, whereas values closer to zero (e.g., $ICC_{DE} = .10$) would suggest that the siblings show divergent patterns of correlations with these variables. Notably, in contrast to other ways to estimate profile agreement (e.g., Pearson *r*), $ICC_{DE}$ takes into account similarity in the absolute value of the correlations as well as the rank-ordering of correlations, both of which are important for understanding the degree to which two constructs are siblings.[2] Hyatt et al. (2018) used this method to quantify profile agreement between the nomological networks of self-esteem and narcissism and found that the profiles were unrelated ($ICC_{DE} = -.05$), suggesting that they do not have similar nomological networks. Hyatt et al. (2018) examined which specific variables were driving the lack of profile similarity between self-esteem and narcissism, and found that antagonism showed the largest discrepancy between narcissism ($r = .50$) and self-esteem ($r = -.36$). An important caveat when comparing nomological networks is that the degree of similarity will depend critically on the particular set of constructs/measures included in the analyses, so it is essential to examine a comprehensive set of constructs that are theoretically relevant to each of the two potential siblings.

### *4. Strong Association Between Measures*

Perhaps most intuitively, sibling similarity can be assessed by examining the strength of the correlation between measures of the constructs. Constructs whose measures are weakly or completely uncorrelated (e.g., $r = .00–.19$) may be considered distant cousins or lacking any familial relationship. At the other end of the continuum, very strong correlations (e.g., $r = .80–1.00$) are often indicative of being identical twins. Sibling constructs typically have measures that correlate correlations between measures somewhere in the middle (e.g., $r = .20–.60$). However, these are by no means hard cut-offs. Instead, the location of the upper and lower thresholds depends on the research area, the degree of measurement error (which can attenuate correlations between measures of sibling constructs), and the possibility that correlations may be inflated or attenuated by shared versus nonshared method variance. Therefore, in addition to using construct valid measures, we also recommend evaluating Criterion 4 by examining correlations between latent variables, which account for measurement error (Borsboom, 2008; Westfall & Yarkoni, 2016).

Measures of narcissism and self-esteem typically correlate in the .30s and .40s (Hyatt et al., 2018; Orth et al., 2016; Trzesniewski et al., 2008). These correlations indicate that they share a moderate amount of variance, but they do not measure the same construct. The association between narcissism and self-esteem is linear (rather than curvilinear) suggesting that it holds across all levels of self-esteem (Crowe et al., 2018). However, the association between narcissism and self-esteem does across different facets of narcissism is examined. For example, Krizan and Herlache (2018) found that higher levels of superiority/grandiosity were associated with higher self-esteem, whereas higher levels of vulnerable narcissism were associated with lower self-esteem, and exploitativeness/entitlement was not significantly associated with self-esteem. Therefore, with multifaceted constructs, it is important to consider which facet of the construct is being measured and how this might influence the findings. It is possible for two constructs to be weakly correlated, but one particular facet of a construct is strongly correlated with (and therefore shares a sibling relationship with) the other construct.[3]

Notably, as with the other criteria presented here, a high correlation between measures is not a necessary condition for identifying sibling constructs. In fact, if researchers rely too heavily on correlations between measures to identify sibling constructs, they risk overlooking potential sibling constructs that do not correlate highly. For example, implicit and explicit prejudice show relatively modest correlations with each other (meta-analytic *r*s ranging from .17 to .25;

Cameron et al., 2012; Hofmann et al., 2005), but there are important conceptual similarities between them (Blair et al., 2005; Dovidio et al., 2001; Wilson et al., 2000). The low correlation may be due to the fact that the measures used to assess them—computer-based reaction times (e.g., Implicit Association Test [IAT] for implicit prejudice; Greenwald et al., 1998) versus self-reports (for explicit prejudice)—are highly saturated with method-specific variance, leaving little substantive variance left to correlate across measures (Dang et al., 2020). The low correlation between implicit and explicitly prejudice may also be influenced by the low test–retest reliability of measures of implicit prejudice (Bosson et al., 2000; Fazio & Olson, 2003; Greenwald & Nosek, 2001). Just as low correlations can be misinterpreted as evidence that two constructs are not siblings, high correlations can also be misinterpreted. For example, measures of anxiety and depression correlate as highly as .70 and, although they share a common parent (negative emotionality), there are important theoretical and neurobiological distinctions between the two and they should not be jingled together (Watson et al., 1988). Thus, despite being common practice, researchers should not rely solely, or even primarily, on concurrent correlations to make inferences about relations between constructs.

Methodological issues often exacerbate problems with using this criterion to identify sibling constructs. In particular, researchers studying related constructs with different methods that show little empirical overlap are left to decide whether they are actually assessing different constructs or whether they are getting at different processes underlying the same construct.[4] For example, questionnaire and task-based measures of self-regulation are assumed to measure a conceptually similar construct (e.g., the ability to regulate impulses and inhibit a dominant response in favor of a behavior that facilitates attainment of long-term goals), but recently researchers have started to argue that they measure distinct processes (e.g., Malanchini et al., 2019). Friedman and Banich (2019) stated that "the psychological constructs assessed by self-regulation questionnaires and behavioral tasks are not interchangeable" (p. 24396), and Enkavi et al. (2019) stated, more bluntly, that "self-report surveys and cognitive tasks measure separate constructs" (p. 24398). Dang et al. (2020) posit multiple reasons why these measures assess distinct response processes, arguing that (a) behavioral measures tap responses in a specific structured situation, whereas self-reports assess behavioral tendencies across a wide range of unstructured situations, and (b) behavioral measures examine maximal performance via reaction time and accuracy, whereas self-reports reflect judgment about one's typical performance. We agree that these constructs are not identical, given the evidence for distinct processes, but go further to view them as having a sibling relationship, given their strong conceptual overlap.

These considerations also highlight the importance of using multimethod assessment of constructs, when possible, so that conclusions about the degree of sibling relatedness are not overly influenced by method factors. To systematically disentangle substantive (i.e., trait) variance and method variance, researchers can use multitrait-multimethod (MTMM) matrices. MTMM matrices include multiple distinct traits (e.g., potential siblings) and multiple methods for assessing each trait (e.g., self-report, informant-report, behavioral measures) and can tease apart conceptual and methodological overlap (D. T. Campbell & Fiske, 1959). For example, if researchers create an MTMM matrix with three measures of Sibling A and three measures of Sibling B, we should expect, as with all MTMM matrices, the heteromethod-monotrait coefficients to have the highest values, followed by the monomethod-heterotrait coefficients, and then the heterotrait-heteromethod coefficients. However, if the two constructs are close siblings, then we would expect stronger values in the monomethod-heterotrait and heterotrait-heteromethod cells.

### *5. Measures Form a Strong General Factor*

Researchers can evaluate the degree to which constructs are siblings using confirmatory factor analysis (CFA). To start, researchers can use items from measures of both constructs to specify a one-factor and a two-factor CFA model (one representing each construct), and assess which model fits better (see Figure 2). If the two-factor model does not fit significantly better than the one-factor model, it suggests they may be twin rather than sibling constructs. Notably, when fitting the two-factor CFA model with sibling constructs, it is likely that at least some items will load onto both factors (Asparouhov & Muthén, 2009); however, it is difficult to determine which items will have high cross-loadings a priori. Because unmodeled cross-loadings can inflate correlations between measures of constructs, this reinforces the importance of using criteria beyond simply calculating correlations (Criterion 4). Researchers can also specify more complex models, such as bifactor models, which indicate how items are related to both general and domain-specific factors. In these models, the general (i.e., parent) factor accounts for the shared variance across items, whereas the domain-specific (i.e., sibling) factors account for the unique influence of each sibling over and above the general factor (Chen et al., 2006; Reise et al., 2010; see Figure 2). Alternatively, a hierarchical CFA model includes a higher order factor with lower order factors for each sibling construct, which separates the unique variance among the siblings at the lower level and also demonstrate that their shared variance is due to the higher order (i.e., parent) factor (see Figure 2). With both bifactor and hierarchical models, a strong general factor would indicate the presence of sibling constructs (or even twins, depending on the strength of the general factor). However, there are numerous concerns with bifactor models in particular, including evidence of overfitting, unstable parameter estimates, and statistical bias in favor of bifactor models, even

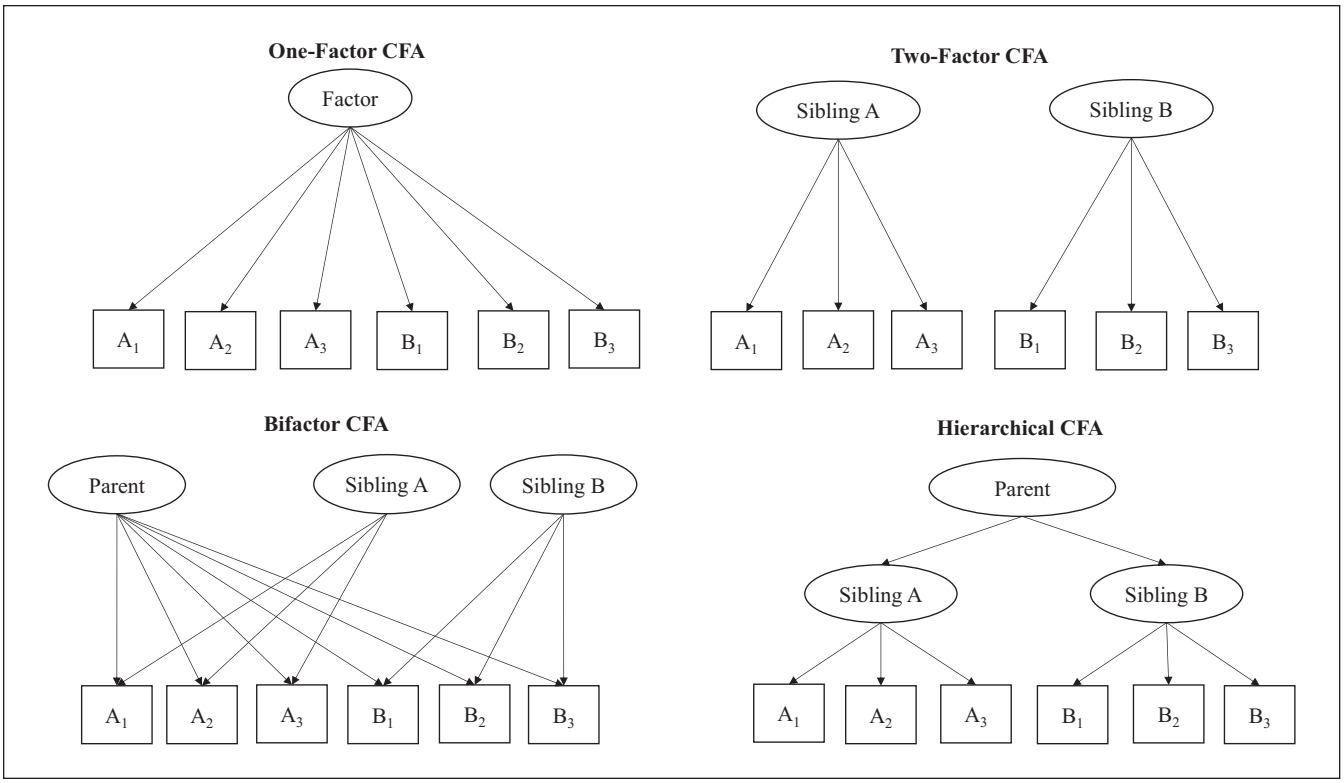


**Figure 2.** Path diagrams of different types of CFAs.
*Note*. A one-factor CFA model fitting best suggests the presence of jangling, whereas a two-factor CFA model fitting best suggests that constructs are distinct. For bifactor and hierarchical CFA models, the strength of the parent factor is an indicator of the degree of the sibling relationship. CFA = confirmatory factor analysis.

when another (simpler) model fits similarly well (Morgan et al., 2015; Murray & Johnson, 2013; Watts et al., 2019). Therefore, it is important to compare competing measurement models and use caution when interpreting the results, especially because mis-specifying the model at the measurement level can result in problems when testing structural models.

Although we do not know of any research that has compared different CFA models of self-esteem and narcissism items, future research could fit one-factor, two-factor, bifactor, and hierarchical models to common measures of self-esteem (e.g., Rosenberg Self-Esteem Scale; Rosenberg, 1965) and narcissism (e.g., Narcissistic Personality Inventory; Raskin & Terry, 1988) and compare evidence for the different factor structures. Given previous empirical work (W. K. Campbell et al., 2002; Hyatt et al., 2018; Trumpeter et al., 2006), we would expect a one-factor model to fit poorly and the bifactor and hierarchical models to fit well. In a bifactor model, we predict that the general factor would represent positive views of the self and the specific sibling factors representing leadership and lack of self-criticism (specific to self-esteem) and exploitativeness and antagonism (specific to narcissism).

## *6. Measures Show Little Incremental Validity Over Each Other*

When measures of constructs are highly correlated, researchers often attempt to demonstrate their unique importance by including both in a prediction model and having them "compete" to explain variance in the outcome. In other words, researchers test whether measures of the sibling constructs show incremental validity over each other (Sechrest, 1963). Taking this multivariate approach and including multiple potential sibling constructs in the same model can isolate their unique associations, which supplements the univariate analyses often used in nomological network analyses (see Edershile et al., 2019 for an example). However, evidence of incremental validity on its own does not ensure that two constructs have not been jangled and measurement issues (e.g., measurement error) can complicate evaluation of this criterion (Wang & Eastwick, 2020). In addition, this criterion depends, in large part, on the chosen dependent variables. For example, sibling constructs may show incremental validity over each other when used to predict certain dependent variables, but not others. Thus, this criterion should always be considered using dependent variables that cover a wide

scope of the constructs' nomological networks (see Criteria 2 and 3).

Previous research examining the incremental validity of self-esteem over narcissism, and vice versa, has led to the realization that they function as mutual suppressor variables. A suppressor variable removes or suppresses criterion-irrelevant variance from another variable; therefore, statistically controlling for a suppressor allows the construct of interest to demonstrate its predicted relation with an outcome (Paulhus et al., 2004). Previous studies have found that including self-esteem and narcissism in a multiple regression model strengthens their associations with theoretically relevant outcomes, including amplifying self-esteem's association with social relationships and mental health and narcissism's association with aggression and antisocial behavior (Locke, 2009; Paulhus et al., 2004; Tracy et al., 2009). Thus, when making claims about either self-esteem or narcissism's effect on some outcome, it is theoretically and empirically important to control for the other variable. However, statistically controlling for another variable raises some interpretational complexities. For example, how should we conceptualize "narcissism free self-esteem" and "self-esteem free narcissism" (Paulhus et al., 2004). Conceptually, self-esteem with narcissistic self-aggrandizement removed is closer to genuine self-esteem, whereas narcissism with self-esteem removed is more like pure self-aggrandizement, not just self-confidence. In this context, statistical control serves to purify the measures, making them conceptually closer to the theorized constructs.

However, in other contexts, statistical control runs the risk of "throwing out the baby with the bathwater." For example, if measures of implicit prejudice are assumed to reflect a person's true underlying prejudicial attitudes, then it would not make sense to examine the effects of explicit prejudice controlling for implicit prejudice because the partialed effects would no longer reflect prejudicial attitudes but rather a person's willingness to explicitly acknowledge prejudicial attitudes. Similarly, if racism and sexism both reflect a belief that "certain groups are 'better' than others" (Sidanius, Pratto, & Bobo, 1994, p. 999), which may be driven by a broad personality factor such as social dominance orientation (Pratto et al., 1994; Sidanius, Pratto, & Mitchell, 1994) or right-wing authoritarianism (Whitley, 1999), then examining the effect of racism after controlling for sexism (or vice versa) would remove a core part of what it means to be a racist or sexist person. Thus, tests of incremental effects raise a number of conceptual and methodological issues, and should only be interpreted after careful consideration of these issues.

Incremental validity, and the other empirical criteria discussed so far can be evaluated using exclusively cross-sectional data. However, the next set of criteria require longitudinal data, which can provide additional insight into the relation between constructs and further elucidate their sibling relationship. In particular, researchers can use longitudinal data to explore whether potential sibling constructs have similar developmental trajectories (Criterion 7), have shared underlying causes (Criterion 8), have a causal effect on each other (Criterion 9), and are trait or state manifestations of the same underlying construct (Criterion 10).

### 7. Similar Developmental Trajectories

Constructs may share a sibling relationship to the extent that they have similar developmental trajectories. That is, sibling constructs might be expected to develop in sync with each other (i.e., codevelop). For example, self-control and externalizing problems (e.g., Oppositional Defiant Disorder, Conduct Disorder) exhibit correlated developmental trajectories across adolescence, such that individuals who show steeper decreases in self-control tend to show steeper increases in externalizing symptoms (Atherton et al., 2020). Indeed, there are potential theoretical explanations underlying this codevelopment, including that it is due to shared etiological factors such as parent antisocial behavior or development of the frontal lobes (Durbin & Hicks, 2014; Krueger et al., 2002). The most common approach to examining the association between developmental trajectories is with bivariate latent growth curve modeling (MacCallum et al., 1997), in which univariate growth curves are specified for each individual construct and then correlations are examined between the slopes of the two constructs (Figure 3). However, if the developmental associations among several constructs are hypothesized to derive from a higher order common "cause" or parent factor, then a more appropriate model is the factor of curves (FOCUS) model (Isiordia et al., 2017). The FOCUS model provides a way to assess the covariation among multiple developmental measures over time (Figure 3). For example, if a researcher had longitudinal data on the Dark Triad constructs (i.e., narcissism, psychopathy, Machiavellianism; Paulhus & Williams, 2002), they could test whether their growth trajectories form a common factor or whether construct-specific growth factors are needed to adequately account for the data.

Prior research examining the normative developmental trajectories of self-esteem and narcissism suggests that the two constructs may codevelop, but in opposing ways (i.e., as one goes up, the other goes down; Brummelman et al., 2016). In particular, on average, self-esteem increases from adolescence to middle adulthood, before decreasing across older adulthood (Orth & Robins, 2014), whereas narcissism peaks during adolescence, then decreases through middle adulthood before increasing slightly in older adulthood (Foster et al., 2003). However, the fact self-esteem and narcissism follow divergent average trajectories does not mean that *individuals* who increase over time in self-esteem will tend to decrease over time in narcissism. Testing this possibility requires longitudinal data in which self-esteem and

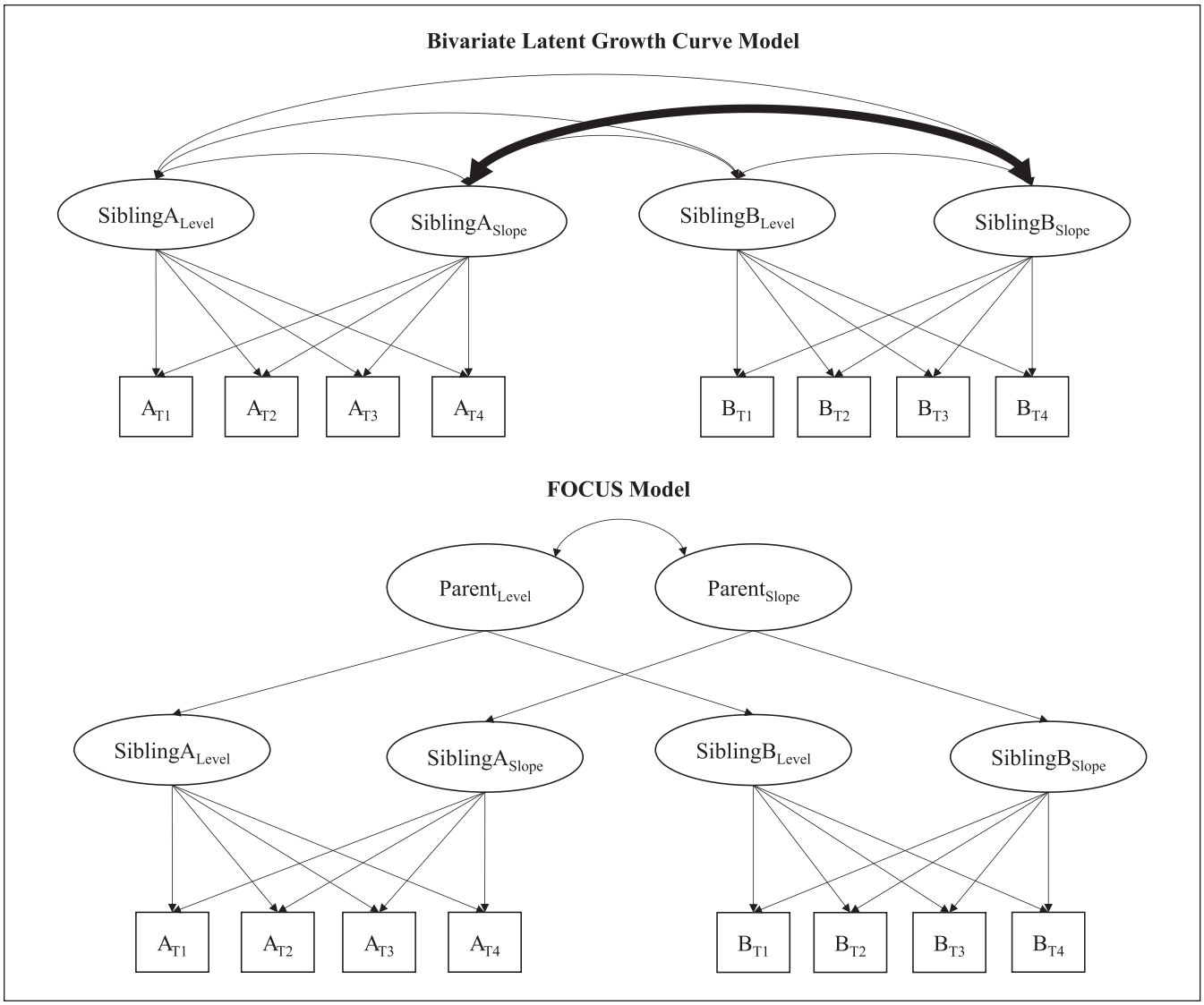


**Figure 3.** Path diagram of bivariate Latent Growth Curve (LGC) and Factor of Curves (FOCUS) models.
*Note*. In a bivariate LGC model, the slope-slope correlation (bolded) depicts the extent to which two constructs codevelop with each other. In the FOCUS model, the higher-order slope factor captures the shared "parent" slope of multiple constructs over time.

narcissism are assessed in the same longitudinal study and individual differences in these trajectories are correlated with each other. Thus, further research is needed to evaluate whether individual trajectories of self-esteem are correlated with individual trajectories of narcissism, or whether there is evidence for a shared growth trajectory (e.g., of positive self-views) that is distinct from specific growth trajectories for self-esteem and narcissism.

## *8. Shared Underlying Causes*

Sibling constructs may be caused by the same underlying factors but develop and manifest differently, suggesting the presence of a common parent. The shared causes that are often examined in psychological research include environmental factors, shared genetic variance, and common underlying neural and physiological mechanisms. For example, some of the sociocontextual factors associated with the development of self-control are also associated with the development of executive function (Atherton et al., 2020). In addition, questionnaire-based measures of self-control and task-based measures of executive functions share considerable genetic variance, despite having little concurrent empirical overlap (Friedman et al., 2020; Malanchini et al., 2019). Similarly, genetic influences explain a substantial amount of variance between self-esteem and negative emotionality (Neiss et al., 2009). Researchers were interested in other shared etiologies between self-control and executive functions could use functional magnetic resonance imaging (fMRI) and other measures of brain activity to determine

whether similar neural activation patterns are associated with self-control and executive function, or between questionnaire-based and task-based measures of self-control.[5]

Research examining the underlying causes of narcissism and self-esteem suggest both similarities and differences. Both constructs emerge during late childhood, when individuals develop the capacity to form stable global self-evaluations and are motivated to maintain favorable self-views (Brummelman et al., 2016; Thomaes et al., 2009) and, like many personality traits, behavioral genetics studies of self-esteem and narcissism, examined separately, show evidence of both genetic heritability and nonshared environmental influences (Luo & Cai, 2018; Neiss et al., 2002; Vernon et al., 2008). However, narcissism may stem from parental coldness, hostility, and unreasonably high expectations, combined with lack of support (Kernberg, 1975; Kohut, 1971; Wetzel & Robins, 2016), whereas supportive parenting and positive social relationships (among other factors) may contribute to the development of self-esteem (Harris & Orth, 2020; Krauss et al., 2020; Orth & Robins, 2014). Indeed, one longitudinal study of children found that parental overevaluation predicted later increases in narcissism, whereas parental warmth predicted later increases in self-esteem (Brummelman et al., 2015). Thus, more research is needed on whether narcissism and self-esteem share common biological causes, but it seems like they diverge from each other with respect to environmental causes. This is an important avenue for future research (Miller et al., in press), and highlights the importance of the sibling construct framework, because "a precise understanding of the distinct roots of narcissism and self-esteem might enable interventions to raise self-esteem while simultaneously curtailing narcissism" (Brummelman et al., 2016, p. 10).

## 9. One Sibling Causes the Other

In contrast to having shared underlying causes, where similar factors cause both of the sibling constructs, there are also situations in which one sibling causes the other(s). These types of causal relations between constructs can be difficult to document, though we will outline a few methods for doing so. The most direct method for determining whether one variable has a causal effect on another is through experimental research or intervention studies. For example, researchers could test whether an experimental manipulation that increases guilt also increases shame, or whether priming implicit prejudice leads to increases in explicit prejudice. However, experimental manipulation is not possible with many constructs that social-personality psychologists study.

Another common approach (albeit controversial; Grosz et al., 2020) to supporting claims about causality is the analysis of prospective effects using data from longitudinal studies. For example, cross-lagged panel models (CLPMs) can be used to estimate prospective, lagged effects of one sibling on another, after taking into account their stability over time and their concurrent association (Biesanz, 2012; Kenny, 1975; Orth et al., 2021; see Figure 4). Such effects help establish the temporal order and direction of effects and thus help rule out alternative (noncausal) explanations. As one example, a large body of longitudinal research using CLPMs is consistent with the claim that low self-esteem leads to depression (Orth et al., 2021; Orth & Robins, 2013). However, traditional CLPMs have been critiqued because they do not distinguish between-person and within-person variance (Berry & Willoughby, 2017; Hamaker et al., 2015; Usami et al., 2019) and various alternative models have also been proposed. Most notably, the random intercept cross-lagged panel model (RI-CLPM) is similar to a CLPM but includes a trait factor that captures stable between-person variance (Hamaker et al., 2015; see Figure 4). The cross-lagged effects in a CLPM indicate the prospective effect of between-person differences in one construct on change in between-person differences in another construct; in contrast, cross-lagged effects in a RI-CLPM estimate the prospective *within-person* effect of deviation from an individual's stable trait level on the *within-person* deviation from the trait level in another construct (Orth et al., 2021).[6] Given that the CLPM and RI-CLPM address different causal questions (Are individuals with low self-esteem more likely to become depressed than individuals with high self-esteem? vs. Do people who have lower self-esteem than usual at one point in time become more depressed than usual at a subsequent time?), researchers may wish to use both CLPMs and RI-CLPMs to examine prospective effects with longitudinal data (Lüdtke & Robitzsch, 2021; Usami et al., 2019). Although there are reasonable concerns about the extent to which these models can actually establish causality, attempts have made to improve causal inference with longitudinal data (Zyphur et al., 2020a, 2020b).

Directed acyclic graphs (DAGs) provide another useful method for identifying causal effects with nonexperimental data, including longitudinal data (Pearl, 2009; Wysocki et al., 2021). A DAG is a figure—similar to a path diagram—that depicts a hypothesized causal model (see Figure 5 for examples and Rohrer, 2018 for details on DAGs in psychology). In DAGs, arrows represent causal effects of one variable on another. The path between two variables, or the way that they are connected by arrows, depicts whether a statistical association will or will not be transmitted, depending on the structure. Given these causal paths, DAGs depict whether a third variable is a confounder, mediator, collider, or other type of third variable (Figure 5). Therefore, DAGs provide information about whether it is appropriate and/or necessary to statistically control for a third variable when trying to calculate the association between a predictor and outcome (see Wysocki et al., 2021 for consequences of statistically controlling for nonconfounders) and allow researchers to estimate the causal effects between variables without experimental data. Because each of these methods for estimating causal associations between sibling constructs require

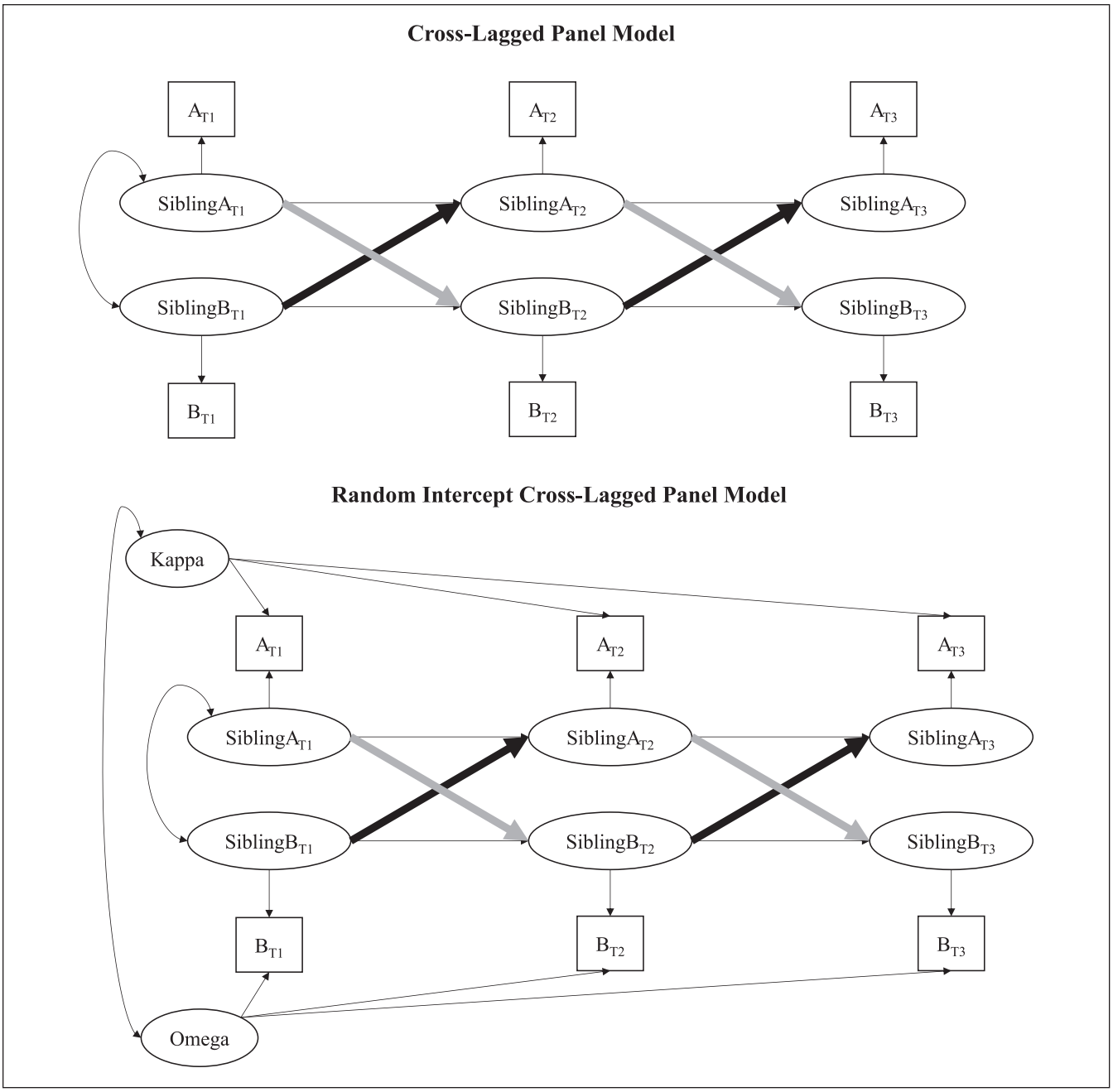


**Figure 4.** Path diagrams of (Random-Intercept) Cross-Lagged Panel Models.
*Note*. The cross-lagged effects (bolded) in the CLPM indicate the prospective effect of between-person differences in one construct on change in between-person differences in another. The cross-lagged effects (bolded) in the RI-CLPM estimate the prospective *within-person* effect of deviation from an individual's stable trait level on the *within-person* deviation from the trait level in another construct (Orth et al., 2021).

substantial theoretical and empirical considerations outlined in the previous criteria, none is perfect for every situation.

Theoretically, narcissistic grandiosity and assertiveness may conceal underlying low self-esteem; in other words, low self-esteem may cause narcissism (Miller et al., 2017). In particular, early childhood experiences arising from parental overidealization and children failing to meet these demands to be perfect may promote defensive, and later unstable, self-esteem (Brummelman et al., 2016; Tracy et al., 2011; Tracy & Robins, 2003). The early humiliation and rejection that underlies unstable self-esteem may then lead to frustration and vengefulness in the face of negative feedback, and these maladaptive tendencies to lash out when criticized generate narcissistic tendencies like exploitativeness and hostility toward others. Conversely, there is reason to believe that narcissistic tendencies may have a causal

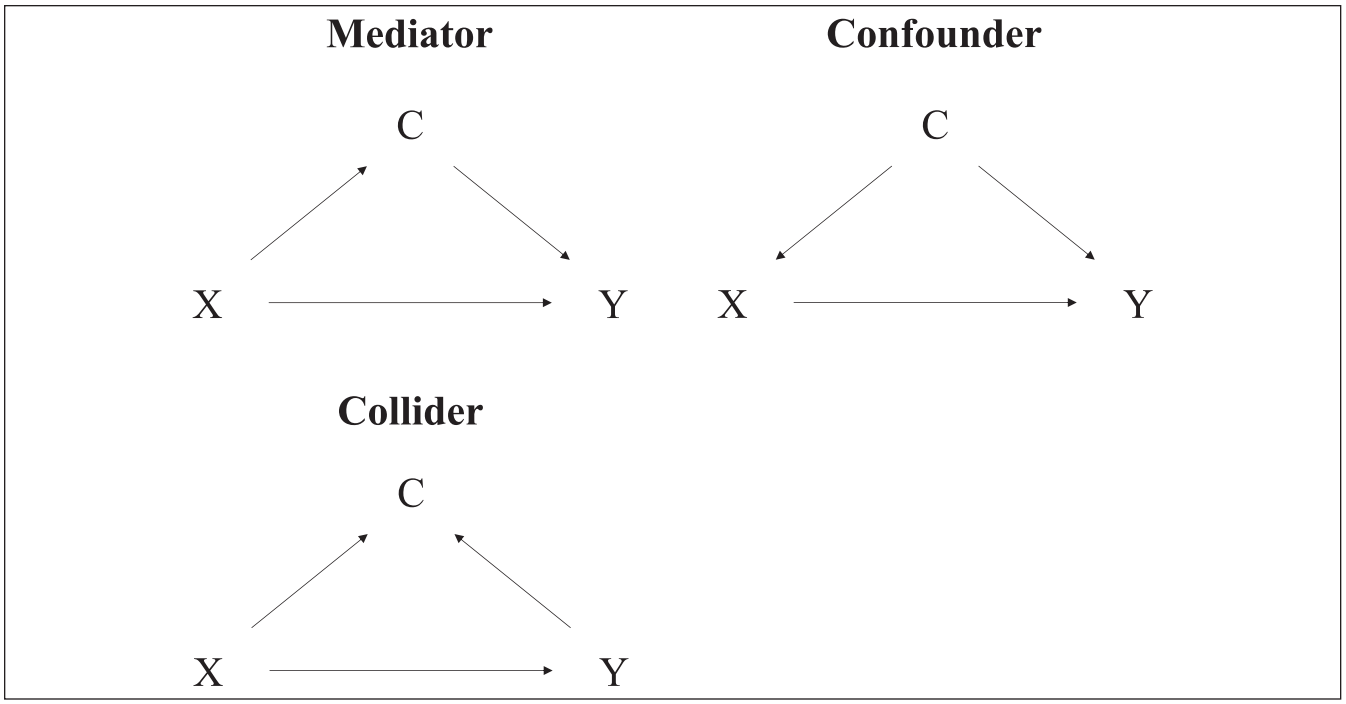


**Figure 5.** Directed acyclic graphs depicting a confounder, mediator, and collider.
*Note.* In all these models, the independent variable is represented with "X," the outcome variable is represented with "Y," and the third variable is represented with "C." A mediator is a variable that is caused by X and causes Y. A confounder is a variable that is a cause of both X and Y. A collider is a variable that is caused by both X and Y. Controlling for a confounder blocks the path that is transmitted through C and unbiases the association between X and Y; however, controlling for a mediator or collider results in biased causal estimates (Rohrer, 2018; Wysocki et al., 2021).

effect on self-esteem. Indeed, individuals with narcissistic tendencies, unlike those with genuine high self-esteem, are often defensive in the face of criticism and this defensiveness may lead to excessive self-aggrandizement, which then leads to higher self-reported self-esteem on explicit measures. To test these theories, several studies have used CLPMs to examine possible causal associations between self-esteem and narcissism and the results have been inconsistent. In a longitudinal study of undergraduate students, narcissism predicted later self-esteem (in the absence of negative life events) but self-esteem did not predict narcissism (Zuckerman & O'Loughlin, 2009). Another study found no prospective effect of self-esteem on narcissism, or vice versa, when stressful life events and depression were included in the models (Orth & Luciano, 2015). Therefore, more research is needed to uncover empirical support for causal relations between self-esteem and narcissism.

### *10. Traits and States*

Finally, researchers should consider the possibility that constructs could share a sibling relationship because they are trait and state manifestations of the same parent construct. For example, negative mood and neuroticism may be, respectively, the state and trait manifestations of negative affect (Rusting & Larsen, 1997). This criterion is not as relevant for sibling constructs that are both considered stable traits (e.g., grit and conscientiousness), but it is useful for demonstrating that two constructs that are closely related according to the other criteria are actually siblings, rather than twins. In addition, this criterion relies heavily on measurement, as measures are typically designed to assess either temporary thoughts, feelings, and behaviors (i.e., states) or more stable and enduring thoughts, feelings, and behaviors (i.e., traits). That is, researchers hoping to assess states might use experience sampling methods (ESM) or instructions about current mood (e.g., "Rate the extent to which you feel sad right now"), whereas those examining traits will use self- or informant-report scales with longer time frames (e.g., "During the past year, how often did you feel sad?"). Given that measures necessarily affect the stability of the constructs being measured, determining whether sibling constructs are trait and state manifestations of the same parent construct can be difficult. One promising method uses intensive longitudinal data to examine whether there are shared nomological associations between measures of traits and state-aggregates (i.e., aggregating numerous repeated state measures over time), or nomological trait-state homomorphy (Rauthmann et al., 2019). Another approach would be to estimate the state and trait variance in each sibling and then test the degree to which the state and trait variance in one construct is associated with the state and trait variance in the other construct (Donnellan et al., 2012).

## Summary

The 10 criteria described above can be used to determine the degree to which constructs share a sibling relationship and what the nature of that relationship is. Rather than relying exclusively on a single criterion to make a yes-or-no decision about sibling status, researchers should consider *all* of the criteria collectively and make a holistic judgment about the *degree* and *nature* of sibling similarity. Indeed, applying the criteria outlined above is the first step toward clarifying associations between the myriad sets of constructs and measures that find themselves lost in the gray area of the jingle-jangle jungle*. Although our approach does not provide a simple yes-or-no decision rule, it does provide common ground so that when researchers debate whether constructs are siblings, they will at least be arguing from a shared understanding of the set of relevant criteria and the specific ways in which each criterion can be evaluated.

For example, after applying each of these criteria to self-esteem and narcissism, we have a clearer understanding of where they overlap (e.g., conceptual similarity due to the shared parent of self-favorability, moderate concurrent correlations) and where they are distinct (e.g., different associations with variables in their nomological networks, measures that show incremental validity over each other, different environmental causes). In addition, we have targeted areas that future research should focus on, including whether there is a general factor of self-esteem and narcissism that emerges in factor analyses, whether and how the constructs codevelop over time, and which, if any, underlying causes they share. Overall, the broad idea that self-esteem and narcissism are closely related, but not jangled, is consistent with previous theoretical and empirical research (e.g., Miller et al., in press); however, the sibling construct framework provides a more formal and systematic analysis than what is often currently used (e.g., examining concurrent correlations) and, in doing so, has important implications for how to dealing with sibling constructs in research.

## What Should You Do With Sibling Constructs?

Despite their prevalence, there is little uniformity in handling sibling constructs across research areas. This is likely because researchers often see them as a nuisance and ignore them altogether. However, proper attention to sibling relationships will improve the quality and robustness of psychological research, and reduce the conceptual fog that envelops areas of research where siblings are prominent. Here, we outline four general recommendations for handling sibling constructs.

### *Apply Our Framework*

First, and most important, we believe that all potential sibling constructs should be evaluated using our 10-criteria framework. Doing so will allow researchers to pinpoint exactly how the constructs are interconnected, which will provide important insights into when interpretational problems might arise in research on the constructs. In some cases, adopting our approach will reveal that the sibling overlap occurs primarily at the conceptual level; in other cases, that it occurs primarily at the measurement level; and, in rare cases, that the two constructs are conceptually and empirically distinct, but they develop and change in much the same way. Furthermore, evaluating constructs using these criteria will reduce the occurrence of jingle-jangle fallacies in social-personality research by explicitly highlighting potential areas of overlap across all 10 criteria. Indeed, as with formal theory, this framework builds upon extant theoretical and empirical work by identifying explicit criteria and providing conceptual and statistical tools (e.g., theory maps and factor analytic models) to interrogate hidden assumptions and evaluate conflicting evidence (Robinaugh et al., 2021). Gray (2017) questioned "How many times has an old construct been 'discovered' under a new name because investigators failed to specify the broader set of connected phenomena?" (p. 738). Our proposed framework provides a way to systematically examine these broader phenomena, and consequently provides a more comprehensive and effective method for identifying sibling constructs and mitigating their adverse effects on research.

The application of our framework will also identify holes in the extant research literature. For example, in some areas of research, a great deal of methodological work has been done to develop empirically distinct measures, yet the conceptual distinction between the constructs remains muddled. In other cases, the exact opposite situation occurs, where the constructs have been clearly conceptually defined as distinct but measures are unable to pick up on their uniqueness. Moreover, few areas of research have systematically applied our proposed development and change criteria, suggesting an important focus for future work.

To promote application of our framework, we encourage researchers writing review articles to explicitly discuss potential siblings of the focal construct and use our framework to evaluate how the focal construct relates to potential siblings both conceptually and empirically. For example, Roberts and colleagues (2014) discuss eight potential siblings of conscientiousness. We also recommend that sibling constructs be acknowledged in constraints on generality (COG) statements (Simons et al., 2017); for example, should findings from a measure of implicit prejudice generalize to a measure of explicit prejudice? If not, why?

*When applying these criteria, researchers should consider the possibility that the findings may differ across developmental periods, cultures, demographic groups, and other facets of generalizability. For example, the sibling relationship between self-esteem and narcissism might be stronger (and/or of a different nature) in individualistic versus collectivistic cultures, and the sibling relationship between implicit and explicit prejudice might be stronger in old age versus young adulthood, given that older individuals have more difficulty inhibiting prepotent responses.

We believe that implementing this framework will mitigate the practice of generating novel-seeming constructs to make one's research program appear unique when there is already an established literature on a closely related construct. In particular, when proposing a new construct, researchers should explicitly consider conceptual and empirical overlap between their "new" construct and any existing constructs in the field. If researchers can only differentiate the constructs by name, and not by providing evidence that they differ with respect to our 10 criteria, then it is an obvious instance of jangling and the new construct/label should be rejected. In this way, our framework puts the onus on researchers proposing new constructs to point at the criteria that differentiate their proposed constructs from existing ones.

Applying this framework to a pair of constructs is not a "one and done" endeavor for researchers. Instead, as noted above, it is more akin to the process of construct validation, which Westen and Rosenthal (2003) described as "not only continuous (a matter of degree, not a categorical distinction between valid and invalid) but continual (a perpetual, self-refining process)" (p. 609).

## *Better to be a Splitter Than a Lumper*

Once sibling constructs have been identified and the nature of their relationship is known, we believe it is generally preferable to be a splitter rather than a lumper. Cronbach (1956) labeled researchers as *splitters* if they tended to highlight fine distinctions among constructs by splitting them into constituent elements, and as *lumpers* if they tended to ignore subtle differences between constructs and view them all through the lens of a broader umbrella construct. When there are *not* important conceptual differences between the sibling constructs, lumping will maximize parsimony and avoid the jangle fallacy. In these cases, it is sometimes appropriate to instead conceptualize a parent construct that ignores any nuance that exists at the sibling level. However, when there are important conceptual distinctions between the constructs, researchers should be careful to split sibling constructs, that is, study them separately. For example, in research on the Dark Triad traits, the recognition that narcissism, Machiallevianism, and psychopathy are not interchangeable, but rather, "a constellation of three conceptually distinct but empirically overlapping personality variables" has helped mitigate problems caused by lumping them together (Furnham et al., 2013, p. 199). Indeed, Paulhus and Williams (2002) encouraged researchers to recognize their conceptual similarities and differences and work toward an integrated understanding of what they share in common and what makes them unique, thus reaping the benefits of both splitting and lumping.

## *Think Before Partialing*

Researchers should avoid mindlessly partialing out shared variance between sibling constructs by statistically controlling for empirical overlap between the measures. Many researchers endorse controlling, which statistically accounts for the existence of a presumed confounder, particularly when sibling constructs are present (Spector & Brannick, 2011). Specifically, researchers often control to "purify" measures of each sibling when they are interested in examining how each sibling, on its own, relates to other variables (as described in Criterion 6 on testing for incremental validity). For example, researchers studying implicit prejudice will often control for explicit prejudice to demonstrate that one has predictive power over the other (e.g., Wilson et al., 2000). Similarly, researchers studying the Dark Triad traits will control for two of the constructs to demonstrate the unique importance of the third (e.g., Jones & Paulhus, 2014).

However, the decision of whether or not to control should involve careful weighing of the pros and cons. Statistical control is a powerful tool, but it can lead to biased estimates if used improperly (Wysocki et al., 2021). Therefore, depending on the nature of the causal relations between sibling constructs (outlined in Criterion 9), controlling for the variance that is shared between measures of the sibling constructs could allow the researcher to understand a psychological phenomenon with greater accuracy, but, if done incorrectly, could undermine the interpretability of their conclusions, especially if the shared variance reflects a psychologically meaningful part of each sibling. That is, if the researcher fails to mitigate the impact of a confounder *and* meaningful shared variance is removed, the researcher is essentially throwing the baby out with the bathwater. Thus, the use of statistical control requires a clear psychological understanding of the causal relations between variables and what the variance to be partialed out represents. Fortunately, by applying the criteria enumerated above, researchers should have a good idea about the areas of conceptual and empirical overlap.

The content of shared variance also depends on the technique used to accomplish statistical control. For example, researchers can calculate a partial correlation, which involves computing the correlation between a focal variable and some outcome, holding constant the sibling for both (Abdi, 2007b). Alternatively, researchers can calculate a part correlation (also called semi-partial correlation; Abdi, 2007b), which involves holding the sibling construct constant only for the focal variable (not the outcome) as in a multiple regression. The aforementioned hierarchical and bifactor models are also useful for understanding and teasing apart variance shared via a parent versus variance unique to each sibling.

However, because any procedure for achieving statistical control necessarily removes substantive variance that is shared between the constructs, statistical control effectively changes the construct that was measured. With siblings, it is essential to understand what this shared variance represents to understand how to interpret the specific variance left over for each sibling. In many cases, removing the shared variance creates a measure of a sibling construct that no longer

maps onto its conceptual definition (or is simply noise). This is the central concern of Lynam and colleagues who discuss the perils of partialing and outline how controlling for related constructs can lead to interpretive difficulties (Lynam et al., 2006; Sleep et al., 2017). Broadly, the problem manifests when researchers analyze residualized scores but draw conclusions about the originally conceptualized construct. Therefore, the authors state that researchers must decide "which construct, the original or residualized, the conclusions are meant to apply" (Lynam et al., 2006, p. 339). This is especially relevant in cases where the sibling constructs are multidimensional and/or have a high degree of empirical overlap (i.e., high concurrent correlations). When sibling constructs have minimal empirical overlap, controlling may not provide any benefits because very little shared variance is being removed. Therefore, researchers must be cognizant of problems related to controlling for sibling constructs and recognize that changes through partialing can have widespread implications for the research findings.

### *Be Transparent*

Given the number of analytic decisions that need to be made around sibling constructs, researchers need to be clear and transparent in reporting relevant decisions about them. Of course, transparency is an important foundation for all psychological research, especially when it comes to measurement (Flake & Fried, 2020; Hussey & Hughes, 2018), but it is even more vital with sibling constructs, given the lack of systematic guidance to constrain researchers' decisions as they traverse through the flourishing garden of forking paths (Gelman & Loken, 2013). These degrees of freedom allow well-intentioned researchers to ignore sibling constructs and/or handle them poorly. For example, they allow researchers to run many different analyses of their sibling constructs (e.g., different approaches to controlling for shared variance) and report only those analyses that achieve significant results. Running multiple analyses that handle sibling constructs differently and then reporting only the significant results constitutes *p*-hacking (Head et al., 2015), and should obviously be avoided.

To alleviate this threat, it is important to create and adhere to reporting norms for handling sibling constructs. Notably, these recommendations are intended to supplement questions aimed at promoting transparency of measurement practices more generally (e.g., What is your construct? How is it being operationalized? Why did you choose that specific measure? Do you have other measures of the same construct?; Flake & Fried, 2020; Hussey & Hughes, 2018). First, researchers should clearly identify any siblings of their focal construct(s), regardless of whether the siblings were measured in the study. Although researchers often vaguely identify sibling constructs in literature reviews, they often fail to include any description of these siblings and/or note theorized overlap in their nomological networks. As noted above, one example of creating a coherent literature in which siblings are studied together is research on the Dark Triad traits; Paulhus and Williams (2002) labeled them as such to encourage researchers to study the three traits in tandem. Indeed, there are claims that the constructs must be studied together because "Only then can their distinctiveness be clarified. If studied alone, any observed correlates may actually reflect overlap with one of the other Dark Triad members" (Jones & Paulhus, 2014, p. 28). Second, if siblings were measured in the study, researchers should always report concurrent correlations between the focal construct and the measures of sibling constructs, ideally using latent variable modeling to reduce the impact of measurement error. Concurrent correlations provide information to the reader about what effects statistical control may have, and the degree to which it is even necessary. Third, researchers should clearly report how sibling constructs were handled analytically. For example, researchers should make clear whether they controlled for sibling constructs and, if so, what impact this might have on interpretation of the results. Researchers may wish to run analyses multiple ways (e.g., with and without controlling for the sibling construct). These types of analyses can be elucidating as long as all results are included in the paper and appropriate statistical considerations are made (e.g., corrections for multiple comparisons; Abdi, 2007a). Furthermore, when examining results from multiple analyses, researchers should consider why there might be serious discrepancies between results, as this may be indicative of other underlying problems (e.g., with measurement). As noted above, in many cases, there is no reason to expect that the effect of Sibling A will hold after controlling for Sibling B, and such a finding should not be interpreted as showing that Sibling A is not associated with the outcome (or dependent variable), but rather that whatever *unique variance is left over* in Sibling A after controlling for its substantive overlap with Sibling B does not predict the outcome. Finally, researchers should preregister all their hypotheses and data analysis plans for sibling constructs to constrain researcher degrees of freedom when analyzing the data and minimize the possibility of selective reporting. Preregistering theoretical claims (i.e., the theoretical rationale for the hypotheses) is particularly important when these theoretical considerations directly affect decisions about statistical analyses.

## Conclusion

As psychologists come to terms with the replicability crisis (Asendorpf et al., 2013; Vazire, 2018), some researchers have identified poorly articulated theories as one of the causes, with efforts to improve theory proposed as one of the solutions (Gray, 2017; Muthukrishna & Henrich, 2019; Navarro, 2021; Szollosi et al., 2019). One launching pad for systematically increasing the theoretical, and in turn, methodological rigor of our work is to explicitly identify and appropriately address sibling constructs.

In an effort to promote the proper handling of sibling constructs, this article began by introducing the idea of sibling constructs and situating them in the literature on jingle-jangle fallacies. Then, we outlined 10 criteria that researchers can apply to critically examine the degree to which constructs share a sibling relationship and the nature of this relationship, using self-esteem and narcissism as a running example. Finally, we presented four broad recommendations for handling sibling constructs. Although problems with sibling constructs are widespread *because* they are so difficult to identify and handle, implementing systematic and effective procedures for identifying and dealing with them will facilitate a more cumulative, collaborative, and robust science of social-personality psychology. We hope that the procedures described here will help social-personality psychologists identify sibling constructs, understand when and why they pose problems for their research, and adopt strategies that ameliorate their adverse influence.

## Acknowledgments

Thank you to Olivia E. Atherton, Wiebke Bleidorn, and Christopher J. Hopwood for comments on an early draft of this article.

## 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 authors were supported by a grant from the National Institute on Aging (R01AG060164) while writing this article.

## ORCID iD

Katherine M. Lawson https://orcid.org/0000-0002-4083-9797

## Notes

1. In this article, we focus on grandiose (rather than the vulnerable) narcissism, unless otherwise noted (Miller et al., 2017, in press).
2. Westen and Rosenthal (2003) described procedures for quantifying the correspondence between the predicted and observed associations in a nomological network, which could be adapted for evaluating sibling constructs by quantifying the correspondence between the observed pattern of associations for Sibling A with the observed pattern of associations for Sibling B.
3. This same issue holds for all the other criteria discussed in this section, both conceptual and empirical.
4. Two constructs might also be correlated because they are causally related, which we address in Criterion 9.
5. A conceptually similar idea would be to examine whether two psychological constructs share the same situational influences or eliciting conditions, which is analogous to etiological factors but over a shorter time period. For example, researchers could test whether the situational factors that lead to impairments in one component of executive function also lead to impairments in other components of executive function.
6. Other models have also been proposed to tease apart between- and within-person lagged effects, but problems arise when fitting these complex models to real data, including model nonconvergence (see Orth et al., 2021).

## References

Abdi, H. (2007a). The Bonferonni and Sidak corrections for multiple comparisons. In N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics* (Vol. 3, pp. 103–107). SAGE.

Abdi, H. (2007b). Part (semi partial) and partial regression coefficients. In N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics* (Vol. 3, pp. 1–9). SAGE.

Arffa, S. (2007). The relationship of intelligence to executive function and non-executive function measures in a sample of average, above average, and gifted youth. *Archives of Clinical Neuropsychology*, *22*(8), 969–978. https://doi.org/10.1016/j.acn.2007.08.001

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J. A., Fiedler, K., Fiedler, S., Funder, D. C., Kliegl, R., Nosek, B. A., Perugini, M., Roberts, B. W., Schmitt, M., van Aken, M. A. G., Weber, H., & Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology: Recommendations for increasing replicability. *European Journal of Personality*, *27*(2), 108–119. https://doi.org/10.1002/per.1919

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(3), 397–438. https://doi.org/10.1080/10705510903008204

Atherton, O. E., Lawson, K. M., Ferrer, E., & Robins, R. W. (2020). The role of effortful control in the development of ADHD, ODD, and CD symptoms. *Journal of Personality and Social Psychology*, *118*, 1226–1246.

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

Baumeister, R. F., Smart, L., & Boden, J. M. (1996). Relation of threatened egotism to violence and aggression: The dark side of high self-esteem. *Psychological Review*, *103*, 5–33.

Berry, D., & Willoughby, M. T. (2017). On the practical interpretability of cross-lagged panel models: Rethinking a developmental workhorse. *Child Development*, *88*(4), 1186–1206. https://doi.org/10.1111/cdev.12660

Biesanz, J. C. (2012). Autoregressive longitudinal models. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 459–471). The Guilford Press.

Blair, C., Zelazo, P. D., & Greenberg, M. T. (2005). The measurement of executive function in early childhood. *Developmental Neuropsychology*, *28*(2), 561–571. https://doi.org/10.1207/s15326942dn2802_1

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, *117*(2), 187–215.

Block, J. (2000). Three tasks for personality psychology. In L. R. Bergman, R. B. Cairns, L. G. Nilsson, & L. Nystedt (Eds.), *Developmental science and the holistic approach* (pp. 155–164). Lawrence Erlbaum.

Borsboom, D. (2008). Latent variable theory. *Measurement: Interdisciplinary Research & Perspective*, *6*(1–2), 25–53. https://doi.org/10.1080/15366360802035497

Bosson, J. K., Swann, W. B., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology*, *79*(4), 631–643. https://doi.org/10.1037/0022-3514.79.4.631

Brummelman, E., Thomaes, S., Nelemans, S. A., Orobio de Castro, B., Overbeek, G., & Bushman, B. J. (2015). Origins of narcissism in children. *Proceedings of the National Academy of Sciences*, *112*(12), 3659–3662. https://doi.org/10.1073/pnas.1420870112

Brummelman, E., Thomaes, S., & Sedikides, C. (2016). Separating narcissism from self-esteem. *Current Directions in Psychological Science*, *25*(1), 8–13. https://doi.org/10.1177/0963721415619737

Burgess, P. W., Alderman, N., Evans, J., Emslie, H., & Wilson, B. A. (1998). The ecological validity of tests of executive function. *Journal of the International Neuropsychological Society*, *4*(6), 547–558. https://doi.org/10.1017/S1355617798466037

Cameron, C. D., Brown-Iannuzzi, J. L., & Payne, B. K. (2012). Sequential priming measures of implicit social cognition: A meta-analysis of associations with behavior and explicit attitudes. *Personality and Social Psychology Review*, *16*(4), 330–350.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105.

Campbell, W. K., Foster, C. A., & Finkel, E. J. (2002). Does self-love lead to love for others?: A story of narcissistic game playing. *Journal of Personality and Social Psychology*, *83*(2), 340–354. https://doi.org/10.1037//0022-3514.83.2.340

Chen, F. F., West, S., & Sousa, K. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, *41*(2), 189–225. https://doi.org/10.1207/s15327906mbr4102_5

Credé, M., Tynan, M. C., & Harms, P. D. (2017). Much ado about grit: A meta-analytic synthesis of the grit literature. *Journal of Personality and Social Psychology*, *113*(3), 492–511. https://doi.org/10.1037/pspp0000102

Cronbach, L. J. (1956). Assessment of individual differences. *Annual Review of Psychology*, *7*(1), 173–196. https://doi.org/10.1146/annurev.ps.07.020156.001133

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302.

Crowe, M. L., Sleep, C. E., Carter, N. T., Campbell, W. K., & Miller, J. D. (2018). Self-esteem and narcissism: An item response theory analysis of curvilinearity. *Personality and Individual Differences*, *128*, 16–20. https://doi.org/10.1016/j.paid.2018.02.016

Dang, J., King, K. M., & Inzlicht, M. (2020). Why are self-report and behavioral measures weakly correlated? *Trends in Cognitive Sciences*, *24*(4), 267–269.

Dickens, L. R., & Robins, R. W. (2020). Pride: A meta-analytic project. *Emotion*. https://doi.org/10.1037/emo0000905

Donnellan, M. B., Kenny, D. A., Trzesniewski, K. H., Lucas, R. E., & Conger, R. D. (2012). Using trait–state models to evaluate the longitudinal consistency of global self-esteem from adolescence to adulthood. *Journal of Research in Personality*, *46*(6), 634–645. https://doi.org/10.1016/j.jrp.2012.07.005

Donnellan, M. B., Trzesniewski, K. H., & Robins, R. W. (2013). Self-esteem: Enduring issues and controversies. In T. Chamorro-Premuzic, S. von Stumm, & A. Furnham (Eds.), *The Wiley-Blackwell handbook of individual differences* (pp. 718–746). Wiley-Blackwell. https://doi.org/10.1002/9781444343120.ch28

Donnellan, M. B., Trzesniewski, K. H., Robins, R. W., Moffitt, T. E., & Caspi, A. (2005). Exploring the link between self-esteem and externalizing behaviors: Low self-esteem is related to antisocial behavior, conduct disorder, and delinquency. *Psychological Science*, *16*, 328–335.

Dovidio, J. F., Kawakami, K., & Beach, K. R. (2001). Implicit and explicit attitudes: Examination of the relationship between measures of intergroup bias. In R. Brown & S. L. Gaertner (Eds.), *Blackwell handbook of social psychology: Intergroup processes* (Vol. 4, pp. 175–197). Wiley.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, *82*(1), 62–68.

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, *92*(6), 1087–1101. https://doi.org/10.1037/0022-3514.92.6.1087

Durbin, C. E., & Hicks, B. M. (2014). Personality and psychopathology: A stagnant field in need of development. *European Journal of Personality*, *28*(4), 362–386. https://doi.org/10.1002/per.1962

Edershile, E. A., Simms, L. J., & Wright, A. G. C. (2019). A multivariate analysis of the Pathological Narcissism Inventory's nomological network. *Assessment*, *26*(4), 619–629.

Enkavi, A. Z., Eisenberg, I. W., Bissett, P. G., Mazza, G. L., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Reply to Friedman and Banich: Right measures for the research question. *Proceedings of the National Academy of Sciences*, *116*(49), 24398–24399. https://doi.org/10.1073/pnas.1917123116

Falkenbach, D. M., Howe, J. R., & Falki, M. (2013). Using self-esteem to disaggregate psychopathy, narcissism, and aggression. *Personality and Individual Differences*, *54*(7), 815–820. https://doi.org/10.1016/j.paid.2012.12.017

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, *54*(1), 297–327. https://doi.org/10.1146/annurev.psych.54.101601.145225

Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, *3*(4), 456–465.

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science*, *8*(4), 370–378. https://doi.org/10.1177/1948550617693063

Foster, J. D., Keith Campbell, W., & Twenge, J. M. (2003). Individual differences in narcissism: Inflated self-views across the lifespan and around the world. *Journal of Research in Personality*, *37*(6), 469–486. https://doi.org/10.1016/S0092-6566(03)00026-6

Friedman, N. P., & Banich, M. T. (2019). Questionnaires and task-based measures assess different aspects of self-regulation: Both are needed. *Proceedings of the National Academy of Sciences*, *116*(49), 24396–24397. https://doi.org/10.1073/pnas.1915315116

Friedman, N. P., Hatoum, A. S., Gustavson, D. E., Corley, R. P., Hewitt, J. K., & Young, S. E. (2020). Executive functions and impulsivity are genetically distinct and independently predict psychopathology: Results from two adult twin studies. *Clinical Psychological Science*, *8*, 519–538. https://doi.org/10.1177/2167702619898814

Furnham, A., Richards, S. C., & Paulhus, D. L. (2013). The dark triad of personality: A 10 year review. *Social and Personality Psychology Compass*, *7*(3), 199–216. https://doi.org/10.1111/spc3.12018

Gelman, A., & Loken, E. (2013). *The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time*. http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf

Gray, K. (2017). How to map theory: Reliable methods are fruitless without rigorous theory. *Perspectives on Psychological Science*, *12*(5), 731–741. https://doi.org/10.1177/1745691617691949

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480.

Greenwald, A. G., & Nosek, B. A. (2001). Health of the Implicit Association Test at age 3. *Zeitschrift für experimentelle Psychologie*, *48*, 85–93.

Grosz, M. P., Rohrer, J. M., & Thoemmes, F. (2020). The taboo against explicit causal inference in nonexperimental psychology. *Perspectives on Psychological Science*, *15*(5), 1243–1255.

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*(1), 102–116. https://doi.org/10.1037/a0038889

Harris, M. A., & Orth, U. (2020). The link between self-esteem and social relationships: A meta-analysis of longitudinal studies. *Journal of Personality and Social Psychology*, *119*, 1459–1477. https://doi.org/10.1037/pspp0000265

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLOS Biology*, *13*(3), Article e1002106. https://doi.org/10.1371/journal.pbio.1002106

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality and Social Psychology Bulletin*, *31*(10), 1369–1385. https://doi.org/10.1177/0146167205275613

Hussey, I., & Hughes, S. (2018). *Hidden invalidity among fifteen commonly used measures in social and personality psychology*. https://doi.org/10.31234/osf.io/7rbfp

Hyatt, C. S., Sleep, C. E., Lamkin, J., Maples-Keller, J. L., Sedikides, C., Campbell, W. K., & Miller, J. D. (2018). Narcissism and self-esteem: A nomological network analysis. *PLOS ONE*, *13*(8), Article e0201088. https://doi.org/10.1371/journal.pone.0201088

Ion, A., Mindu, A., & Gorbănescu, A. (2017). Grit in the workplace: Hype or ripe? *Personality and Individual Differences*, *111*, 163–168. https://doi.org/10.1016/j.paid.2017.02.012

Isiordia, M., Conger, R., Robins, R. W., & Ferrer, E. (2017). Using the factor of curves model to evaluate associations among multiple family constructs over time. *Journal of Family Psychology*, *31*, 1017–1028. http://dx.doi.org/10.1037/fam0000379

Jones, D. N., & Paulhus, D. L. (2014). Introducing the Short Dark Triad (SD3): A brief measure of dark personality traits. *Assessment*, *21*(1), 28–41.

Kelley, T. L. (1927). *Interpretation of educational measurements*. World Book.

Kenny, D. A. (1975). Cross-lagged panel correlation: A test for spuriousness. *Psychological Bulletin*, *82*(6), 887–903. https://doi.org/10.1037/0033-2909.82.6.887

Kernberg, O. F. (1975). *Borderline conditions and pathological narcissism*. Jason Aronson.

Kohut, H. (1971). *The analysis of the self: A systematic approach to the psychoanalytic treatment of narcissistic personality disorders*. University of Chicago Press.

Kramer, A. F., Humphrey, D. G., Larish, J. F., & Logan, G. D. (1994). Aging and inhibition: Beyond a unitary view of inhibitory processing in attention. *Psychology and Aging*, *9*, 491–512.

Krauss, S., Orth, U., & Robins, R. W. (2020). Family environment and self-esteem development: A longitudinal study from age 10 to 16. *Journal of Personality and Social Psychology*, *119*, 457–478.

Krizan, Z., & Herlache, A. D. (2018). The narcissism spectrum model: A synthetic view of narcissistic personality. *Personality and Social Psychology Review*, *22*(1), 3–31. https://doi.org/10.1177/1088868316685018

Krueger, R. F., Hicks, B. M., Patrick, C. J., Carlson, S. R., Iacono, W. G., & McGue, M. (2002). Etiologic connections among substance dependence, antisocial behavior and personality: Modeling the externalizing spectrum. *Journal of Abnormal Psychology*, *111*(3), 411–424. https://doi.org/10.1037//0021-843X.111.3.411

Lilienfeld, S. O., & Strother, A. N. (2020). Psychological measurement and the replication crisis: Four sacred cows. *Canadian Psychology / Psychologie Canadienne*, *61*(4), 281–288. https://doi.org/10.1037/cap0000236

Locke, K. D. (2009). Aggression, narcissism, self-esteem, and the attribution of desirable and humanizing traits to self versus others. *Journal of Research in Personality*, *43*(1), 99–102. https://doi.org/10.1016/j.jrp.2008.10.003

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, *3*, 635–694.

Lüdtke, O., & Robitzsch, A. (2021). *A critique of the random intercept cross-lagged panel model*. https://psyarxiv.com/6f85c/

Luo, Y. L. L., & Cai, H. (2018). The etiology of narcissism: A review of behavioral genetic studies. In A. D. Hermann, A. B. Brunell, & J. D. Foster (Eds.), *Handbook of Trait Narcissism* (pp. 149–156). Springer International Publishing. https://doi.org/10.1007/978-3-319-92171-6_16

Lynam, D. R., Hoyle, R. H., & Newman, J. P. (2006). The perils of partialling: Cautionary tales from aggression and psychopathy. *Assessment*, *13*(3), 328–341. https://doi.org/10.1177/1073191106290562

MacCallum, R. C., Kim, C., Malarkey, W. B., & Kiecolt-Glaser, J. K. (1997). Studying multivariate change using multilevel models and latent curve models. *Multivariate Behavioral Research*, *32*(3), 215–253. https://doi.org/10.1207/s15327906mbr3203_1

Malanchini, M., Engelhardt, L. E., Grotzinger, A. D., Harden, K. P., & Tucker-Drob, E. M. (2019). "Same but different": Associations between multiple aspects of self-regulation, cognition, and academic abilities. *Journal of Personality and*

*Social Psychology*, *117*(6), 1164–1188. https://doi.org/10.1037/pspp0000224

Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Arens, A. K. (2019). The murky distinction between self-concept and self-efficacy: Beware of lurking jingle jangle fallacies. *Journal of Educational Psychology*, *111*(2), 221–353.

McCrae, R. R. (2008). A note on some measures of profile agreement. *Journal of Personality Assessment*, *90*(2), 105–109. https://doi.org/10.1080/00223890701845104

Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science*, *34*(2), 103–115. https://doi.org/10.1086/288135

Miller, J. D., Back, M. D., Lynam, D. R., & Wright, A. G. C. (in press). Narcissism today: What we know and what we need to learn. *Current Directions in Psychological Science*.

Miller, J. D., Lynam, D. R., Hyatt, C. S., & Campbell, W. K. (2017). Controversies in narcissism. *Annual Review of Clinical Psychology*, *13*(1), 291–315. https://doi.org/10.1146/annurev-clinpsy-032816-045244

Millner, A. J., Robinaugh, D. J., & Nock, M. K. (2020). Advancing the understanding of suicide: The need for formal theory and rigorous descriptive research. *Trends in Cognitive Sciences*, *24*(9), 704–716. https://doi.org/10.1016/j.tics.2020.06.007

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100.

Morgan, G., Hodge, K., Wells, K., & Watkins, M. (2015). Are fit indices biased in favor of bi-factor models in cognitive ability research?: A comparison of fit in correlated factors, higher-order, and bi-factor models via Monte Carlo simulations. *Journal of Intelligence*, *3*(1), 2–20. https://doi.org/10.3390/jintelligence3010002

Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence*, *41*(5), 407–422. https://doi.org/10.1016/j.intell.2013.06.004

Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behavior*, *3*, 221–229. https://doi.org/10.1038/s41562-018-0522-1

Navarro, D. J. (2021). *If mathematical psychology did not exist we might need to invent it: A comment on theory building in psychology*. https://psyarxiv.com/ygbjp

Neiss, M. B., Sedikides, C., & Stevenson, J. (2002). Self-esteem: A behavioural genetic perspective. *European Journal of Personality*, *16*(5), 351–367. https://doi.org/10.1002/per.456

Neiss, M. B., Stevenson, J., Legrand, L. N., Iacono, W. G., & Sedikides, C. (2009). Self-esteem, negative emotionality, and depression as a common temperamental core: A study of mid-adolescent twin girls. *Journal of Personality*, *77*(2), 327–346. https://doi.org/10.1111/j.1467-6494.2008.00549.x

Nigg, J. T., Blaskey, L. G., Huang-Pollack, C., & Rappley, M. D. (2002). Neuropsychological executive functions and DSM-IV ADHD subtypes. *Journal of the American Academy of Child and Adolescent Psychiatry*, *41*, 59–66.

Orth, U., Clark, D. A., Donnellan, M. B., & Robins, R. W. (2021). Testing prospective effects in longitudinal research: Comparing seven competing cross-lagged models. *Journal of Personality and Social Psychology*, *120*, 1013–1034. https://doi.org/10.1037/pspp0000358

Orth, U., & Luciano, E. C. (2015). Self-esteem, narcissism, and stressful life events: Testing for selection and socialization. *Journal of Personality and Social Psychology*, *109*(4), 707–721. https://doi.org/10.1037/pspp0000049

Orth, U., & Robins, R. W. (2013). Understanding the link between low self-esteem and depression. *Current Directions in Psychological Science*, *22*(6), 455–460. https://doi.org/10.1177/0963721413492763

Orth, U., & Robins, R. W. (2014). The development of self-esteem. *Current Directions in Psychological Science*, *23*(5), 381–387. https://doi.org/10.1177/0963721414547414

Orth, U., Robins, R. W., Meier, L. L., & Conger, R. D. (2016). Refining the vulnerability model of low self-esteem and depression: Disentangling the effects of genuine self-esteem and narcissism. *Journal of Personality and Social Psychology*, *110*(1), 133–149. https://doi.org/10.1037/pspp0000038

Orth, U., & Robins, R. W. (in press). Is high self-esteem beneficial? Revising a classic question. *American Psychologist*.

Paulhus, D. L., Robins, R. W., Trzesniewski, K. H., & Tracy, J. L. (2004). Two replicable suppressor situations in personality research. *Multivariate Behavioral Research*, *39*(2), 303–328. https://doi.org/10.1207/s15327906mbr3902_7

Paulhus, D. L., & Williams, K. M. (2002). The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, *36*(6), 556–563. https://doi.org/10.1016/S0092-6566(02)00505-6

Pearl, J. (2009). *Causality*. Cambridge University Press.

Peck, S. C. (2004). *The jingle-jangle jungle of identity: Historical perspective and future directions*. Unpublished manuscript, University of Michigan, Ann Arbor.

Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology*, *67*(4), 741–763. https://doi.org/10.1037/0022-3514.67.4.741

Rauthmann, J. F., Horstmann, K. T., & Sherman, R. A. (2019). Do self-reported traits and aggregated states capture the same thing? A nomological perspective on trait-state homomorphy. *Social Psychological and Personality Science*, *10*(5), 596–611. https://doi.org/10.1177/1948550618774772

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, *92*(6), 544–559. https://doi.org/10.1080/00223891.2010.496477

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.

Reschly, A. L., & Christenson, S. (2012). Jingle, jangle, and conceptual haziness: Evolution and future directions of the engagement constructs. In S. L. Christenson, A. L. Rechly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 3–19). Springer.

Roberts, B. W., Lejuez, C., Krueger, R. F., Richards, J. M., & Hill, P. L. (2014). What is conscientiousness and how can it be assessed? *Developmental Psychology*, *50*(5), 1315–1330. https://doi.org/10.1037/a0031109

Robinaugh, D. J., Haslbeck, J. M. B., Ryan, O., Fried, E. I., & Waldorp, L. J. (2021). Invisible hands and fine calipers: A call to use formal theory as a toolkit for theory construction. *Perspectives on Psychological Science*, *16*, 725–743.

Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, *1*(1), 27–42.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Rosenberg, M. (1989). *Society and the adolescent self-image* (Rev. ed.). Wesleyan University Press.

Rusting, C. L., & Larsen, R. J. (1997). Extraversion, neuroticism, and susceptibility to positive and negative affect: A test of two theoretical models. *Personality and Individual Differences*, *22*(5), 607–612. https://doi.org/10.1016/S0191-8869(96)00246-2

Sechrest, L. (1963). Incremental validity: A recommendation. *Educational and Psychological Measurement*, *23*(1), 153–158.

Sidanius, J., Pratto, F., & Bobo, L. (1994). Social dominance orientation and the political psychology of gender: A case of invariance? *Journal of Personality and Social Psychology*, *67*(6), 998–1011.

Sidanius, J., Pratto, F., & Mitchell, M. (1994). In-group identification, social dominance orientation, and differential intergroup social allocation. *The Journal of Social Psychology*, *134*(2), 151–167. https://doi.org/10.1080/00224545.1994.9711378

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on Generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, *12*(6), 1123–1128.

Sleep, C. E., Lynam, D. R., Hyatt, C. S., & Miller, J. D. (2017). Perils of partialing redux: The case of the Dark Triad. *Journal of Abnormal Psychology*, *126*(7), 939–950. https://doi.org/10.1037/abn0000278

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, *113*(1), 117–143. https://doi.org/10.1037/pspp0000096

Spector, P. E., & Brannick, M. T. (2011). Methodological urban legends: The misuse of statistical control variables. *Organizational Research Methods*, *14*(2), 287–305.

Swim, J. K., Aikin, K. J., Hall, W. S., & Hunter, B. A. (1995). Sexism and racism: Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology*, *68*(2), 199–214. https://doi.org/10.1037/0022-3514.68.2.199

Szollosi, A., Kellen, D., Navarro, D., Shiffrin, R., van Rooij, I., Van Zandt, T., & Donkin, C. (2019). *Is preregistration worthwhile?* https://doi.org/10.31234/osf.io/x36pz

Tangey, J., & Dearing, R. L. (2002). *Shame and guilt*. The Guilford Press.

Tangey, J. P., Miller, R. S., Flicker, L., & Barlow, D. H. (1996). Are shame, guilt, and embarrassment distinct emotions? *Journal of Personality and Social Psychology*, *70*(6), 1265–1269.

Tangney, J. P., Wagner, P. E., & Gramzow, R. (1989). *The Test of the Self-Conscious Affect (TOSCA)*. George Mason University.

Thomaes, S., Bushman, B. J., Orobio de Castro, B., & Stegge, H. (2009). What makes narcissists bloom? A framework for research on the etiology and development of narcissism. *Development and Psychopathology*, *21*(4), 1233–1247. https://doi.org/10.1017/S0954579409990137

Thorndike, E. L. (1904). *An Introduction to the theory of mental and social measurements*. Science Press.

Tracy, J. L., Cheng, J. T., Robins, R. W., & Trzesniewski, K. H. (2009). Authentic and hubristic pride: The affective core of self-esteem and narcissism. *Self and Identity*, *8*(23), 196–213. DOI: 10.1080/15298860802505053

Tracy, J. L., & Robins, R. W. (2003). "Death of a (narcissistic) salesman:" An integrative model of fragile self-esteem. *Psychological Inquiry*, *14*(1), 57–62.

Tracy, J. L., & Robins, R. W. (2006). Appraisal antecedents of shame and guilt: Support for a theoretical model. *Personality and Social Psychology Bulletin*, *32*(10), 1339–1351. https://doi.org/10.1177/0146167206290212

Tracy, J. L., & Robins, R. W. (2007). Self-conscious emotions: Where self and emotion meet. In C. Sedikides & S. J. Spencer (Eds.), *The self* (pp. 187–209). Psychology Press.

Tracy, J. L., Robins, R. W., Schriber, R. A., & Solomon, M. (2011). Is emotion recognition impaired in individuals with autism spectrum disorders? *Journal of Autism and Developmental Disorders*, *41*(1), 102–109. https://doi.org/10.1007/s10803-010-1030-y

Trzesniewski, K. H., Donnellan, M. B., & Robins, R. W. (2008). Is generation me really more narcissistic than previous generations? *Journal of Personality*, *76*(4), 903–918. https://doi.org/10.1111/j.1467-6494.2008.00508.x

Trumpeter, N., Watson, P. J., & O'Leary, B. J. (2006). Factors within Multidimensional Perfectionism Scales: Complexity of relationships with Self-Esteem, Narcissism, Self-Control, and Self-Criticism. *Personality and Individual Differences*, *41*(5), 849–860. https://doi.org/10.1016/j.paid.2006.03.014

Usami, S., Murayama, K., & Hamaker, E. L. (2019). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods*, *24*, 637–657. https://doi.org/10.1037/met0000210

Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and progress. *Perspectives on Psychological Science*, *13*(4), 411–417. https://doi.org/10.1177/1745691617751884

Vernon, P. A., Villani, V. C., Vickers, L. C., & Harris, J. A. (2008). A behavioral genetic investigation of the Dark Triad and the Big 5. *Personality and Individual Differences*, *44*(2), 445–452. https://doi.org/10.1016/j.paid.2007.09.007

Wang, Y. A., & Eastwick, P. W. (2020). Solutions to the problems of incremental validity testing in relationship science. *Personal Relationships*, *27*(1), 156–175. https://doi.org/10.1111/pere.12309

Watson, D., Clark, L. A., & Carey, G. (1988). Positive and negative affectivity and their relation to anxiety and depressive disorders. *Journal of Abnormal Psychology*, *97*, 363–353.

Watts, A. L., Poore, H. E., & Waldman, I. D. (2019). Riskier tests of the validity of the bifactor model of psychopathology. *Clinical Psychological Science*, *7*, 1285–1303.

Weidman, A. C., Steckler, C. M., & Tracy, J. L. (2017). The jingle and jangle of emotion assessment: Imprecise measurement, casual scale usage, and conceptual fuzziness in emotion research. *Emotion*, *17*(2), 267–295.

Westen, D., & Rosenthal, R. (2003). Quantifying construct validity: Two simple measures. *Journal of Personality and Social Psychology*, *84*(3), 608–618. https://doi.org/10.1037/0022-3514.84.3.608

Westfall, J., & Yarkoni, T. (2016). Statistically controlling for confounding constructs is harder than you think. *PLOS ONE*,

*11*(3), Article e0152719. https://doi.org/10.1371/journal.pone.0152719

Wetzel, E. & Robins, R. W. (2016). Are parenting practices associated with the development of narcissism? Findings from a longitudinal study of Mexican-origin youth. *Journal of Research in Personality*, *63*, 84–94.

Whitley, B. E. (1999). Right-wing authoritarianism, social dominance orientation, and prejudice. *Journal of Personality and Social Psychology*, *77*(1), 126–134.

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, *107*(1), 101–126. https://doi.org/10.1037/0033-295X.107.1.101

Wysocki, A., Lawson, K. M., & Rhemtulla, M. (2021). *Statistical control requires causal justification*. Manuscript under review.

Zuckerman, M., & O'Loughlin, R. E. (2009). Narcissism and well-being: A longitudinal perspective. *European Journal of Social Psychology*, *39*(6), 957–972. https://doi.org/10.1002/ejsp.594

Zyphur, M. J., Allison, P. D., Tay, L., Voelkle, M. C., Preacher, K. J., Zhang, Z., Hamaker, E. L., Shamsollahi, A., Pierides, D. C., Koval, P., & Diener, E. (2020a). From data to causes I: Building A general cross-lagged panel model (GCLM). *Organizational Research Methods*, *23*(4), 651–687. https://doi.org/10.1177/1094428119847278

Zyphur, M. J., Voelkle, M. C., Tay, L., Allison, P. D., Preacher, K. J., Zhang, Z., Hamaker, E. L., Shamsollahi, A., Pierides, D. C., Koval, P., & Diener, E. (2020b). From data to causes II: Comparing approaches to panel data analysis. *Organizational Research Methods*, *23*, 688–716. https://doi.org/10.1177/1094428119847280