


Questioning Psychological Constructs: Current Issues and Proposed Changes

Paul De Boeck^a , Jolynn Pek^a, Katherine Walton^a, Duane T. Wegener^a, Brandon M. Turner^a, Barbara L. Andersen^a, Theodore P. Beauchaine^b, Luc Lecavalier^a, Jay I. Myung^a, and Richard E. Petty^a

^aDepartment of Psychology, The Ohio State University, Columbus, Ohio; ^bDepartment of Psychology, University of Notre Dame, Notre Dame, Indiana

ABSTRACT

Constructs are central to psychology. We describe two current trends as responses to dissatisfactions with the abstract nature of constructs and with uncertain and variable research findings: a trend away from constructs toward specific notions and effects, and sharper construct definitions with improved construct measurement. We explain that the issues in psychology reflect the complexity and variable nature of psychological phenomena. Rather than following the trends, we propose a reformulation of the construct notion to accommodate complexity and uncertainty. To provide background, we describe the historical development and epistemology of constructs, and highlight the explanatory and integrative role of constructs. Our reformulation implies that constructs are (1) composite, (2) organized in a hierarchy with overlap, (3) variable, with heterogeneous measurement results. Methods can be and are being developed accordingly. We close with considerations for further debate.

KEYWORDS

constructs, variability, uncertainty, measurement, replication



The notion of constructs is fundamental to theory and research in psychology and its allied disciplines. Constructs are defined operationally in empirical studies and findings are interpreted in terms of constructs. Despite the fundamental role constructs serve in research, the notion of constructs has received minimal attention in the bulk of contemporary literature. Slaney's (2017) discussion is an exception. We observe a silent trend in much of the literature characterized by moving *away* from constructs in psychological science with little explicit criticism of their value. This move away implies a *downscaling of constructs* favoring observables and notions closer to observables over unobservable abstract constructs. Those moving away from constructs seek greater specificity by staying closer to observables (e.g., Borsboom et al., 2021) and specific effects (Open Science Collaboration, 2015). This trend arises from dissatisfactions with the vagueness of constructs, variable results, lack of progress in accumulating knowledge, and numerous calls for direct rather than conceptual replications in addressing failures to replicate.


Conversely, in some corners of the literature, explicit calls to move *toward upscaling operational definitions* (ODs) for better measurement of constructs have appeared. Most of these calls for better measurement acknowledge the problems noted above, and inspired by dissatisfactions with these problems, they recommend tightened definitions of constructs and improved measurement (e.g., Flake & Fried, 2020).

Our position is neither of the two. We accept the issues outlined above as sensible reflections of *uncertainty* in

psychological science—an essential feature of most psychological phenomena, mainly due to various types of complexities of human behavior. In response, and for the sake of discussion, we are setting up a controversy. What is to blame for these issues, the research practices or reality? We choose the latter and call for an approach and methods that accord with this reality. Koch (1992) explains that dissatisfactions are inherent to studying a complex reality, that the frustrations repeat from time-to-time in psychological science and are followed by efforts to improve methods and to define univocal concepts. Rather than directly acting on the frustrations we believe that construct abstractness and related heterogeneity of content reflect the true complexity of psychological phenomena that leads to the frustration, and that accepting complexity-based uncertainty and devising better approaches to characterize and assess it are needed to improve our inferences. With this perspective in mind, we present a reformulated notion of constructs as areas to represent the composite (i.e., not unitary) content of constructs. Treating constructs as being composite in nature better reflects psychological reality, which is often complex and necessarily uncertain. Constructs that better reflect reality cannot be captured with unitary notions that, like points, do not allow for compositeness.

To contextualize our reformulation of constructs, we begin in Section “Constructs and Operational Definitions as Commonly Used in Psychology” by identifying the fundamental problem of the gap between constructs and their

CONTACT Paul De Boeck  deboeck.2@osu.edu  Department of Psychology, Ohio State University, Lazenby Hall, 1827 Neil Avenue, Columbus, OH 43210, USA.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/1047840X.2023.2274429>.

© 2024 The Author(s). Published with license by Taylor & Francis Group, LLC

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

ODs. We argue that the ODs necessarily cover only parts of constructs and in a variable way, which results in apparently non-robust and imprecise findings. Motivations to close this gap manifest in the push for novel treatments of constructs (i.e., downscaling constructs or upscaling ODs), whereas acceptance of this gap can be reinforced by increased methodological rigor. Section “Recent Trends” provides more detail and examples of downscaling constructs and upscaling ODs, further explicating why these approaches try to reduce the gap between constructs and their ODs. In Section “History and Epistemology of the Construct Notion,” we briefly review the history and epistemology of constructs to highlight that the construct-OD gap is inherent to the nature of constructs and has been experienced as an issue throughout the history of psychology and in epistemological views. If downscaling constructs, upscaling ODs, or accepting an implicit or explicit ideal of constructs as unitary would yield increased certainty about findings, it might be at the cost of an integration of findings; it might also, instead, lead to an even more scattered field. In contrast, retaining the abstraction borne by constructs and allowing for more uncertainty and variation in the notion offers a path forward to identify common general principles that might facilitate integration across theories and contexts of application.

With the intent of reformulating the notion of constructs to address concerns about downstream sensitivity in results, Section “The Role of Constructs” identifies the purposes of constructs to ensure that our reformulation preserves the utility of constructs in psychological theory and practice. Specifically, we consider three distinct prototypical constructs (attitudes, autism, and evidence accumulation), and examine them for their commonality as constructs. From this exercise, we assert that constructs are abstractions that have integrative and explanatory roles in theory and practice, while not excluding prediction and description roles. In Section “Construct-Related Issues and Possible Solutions,” we provide a reformulation of constructs that maintain their intended purposes while explicitly embracing their abstract, heterogeneous, and composite nature as opposed to constructs that are necessarily unitary, homogeneous, and precise. We also specify the composite nature as hierarchically structured and variable.

Constructs and Operational Definitions as Commonly Used in Psychology

As used in psychological research, the term “construct” is often assumed to be so basic and so broadly understood that it is rarely defined. When the notion of constructs is defined, it is articulated in a rather vague manner, such as “a conceptual term used to describe a phenomenon of interest” (Edwards & Bagozzi, 2000, p. 157) or “some postulated attribute of people” (Cronbach & Meehl, 1955, p. 283; henceforth C&M). Likewise, the notion of constructs is not explicated in the influential Shadish et al. (2002) volume on experimental and quasi-experimental designs. Specific constructs used in theory are often defined more explicitly (e.g.,

attitude, <https://dictionary.apa.org/attitude>; Petty & Cacioppo, 1986), but consensus is rarely achieved, even for ubiquitous constructs such as intelligence (Sternberg & Detterman, 1986) and emotion (Barrett, 2017a, 2017b).

Although we use the term OD for any operationally induced or recorded observable (e.g., presenting a test, conducting an experimental manipulation), C&M were more reluctant because they understood ODs as the operations themselves and not as variation induced or measured, for example, food deprivation to induce hunger (Tolman, 1936). They preferred “measure” and “test” as more directly and substantively connected with constructs. In the view of Chang (2021), the C&M article is a second-generation operationalist approach because the meaning of a concept is broader than its measurement (Chang does not use the term “construct” but “concept” instead). Thus, identifying a construct with just its measurement is a stricter form of operationalism. ODs as used in psychology are based on observable indicators of constructs (Feest, 2005), which is somewhat different from the notion as introduced and discussed by Bridgman (1927). In C&M’s view and in the psychometric literature, ODs are measurement instruments. In his recent review, Tal (2020) connects measurement with ODs of quantity concepts. In a discipline that uses quantitative approaches where possible, it is unsurprising that scores from measurement instruments (tests, and various types of scales) are used as ODs and indicators of constructs. In experimental studies, the ODs of independent variables are the experimental manipulations.

C&M provide the prevailing approach to constructs in psychology, with two important features:

1. *Nodes in a network.* Constructs are nodes in a network with other constructs and with observables. We refer to this as a componential view on constructs, with each construct consisting of one or more components and external components or constructs as nodes in the network. In the psychometric literature, components internal to the construct manifest as multidimensionality and/or subscales whereas components external to the construct manifest as other variables (i.e., constructs or observables). Construct validity is formulated in terms of relations of measures with other measures for the same construct or with measures of related constructs in the nomological network. Construct validity is typically ascertained by observing convergence of scores from a measure with other scores from measures of the same construct.
2. *Not fully captured.* Constructs cannot be fully captured by observables. The full meaning of a construct is broader than its observables and would require that the whole structure be known, which is practically untenable. We refer to this second feature as the *gap between constructs and their ODs* (the construct-OD gap from above). For psychological tests and test scores (or measurement tools in the broader sense), “Construct validation is involved whenever a test is to be interpreted as a measure of some attribute or quality which is not ‘operationally defined.’ The problem faced by investigators is, ‘what constructs account for variance in test

performance?” (C&M, p. 282). That less than 100% variance of test performance is explained (i.e., after a correction for less than perfect reliability) indicates that a construct is not fully captured, and that there must be remaining aspects of the construct that are not reflected in any one measure.

The C&M article focused on psychometrics and individual-difference constructs, but in Cronbach’s later work (e.g., Cronbach et al., 1963), he also described a generalizability framework for instances of implementing and measuring a construct. Thinking of ODs in terms of generalizability agrees with the notion of content validity in C&M. Content validity concerns how well a measure (e.g., the set of items in a test) reflects the content of the construct. This generalizability framework inspired Shadish et al. (2002) to define construct validity in contexts of experimental research as generalizability of inferences from specific experimental manipulations. The Shadish et al. (2002) definition of construct validity implies that constructs cannot be fully captured in an experiment. Cronbach himself (1982) was pessimistic about successful generalization of experimental and quasi-experimental research on educational program evaluations (the focus of his work at the time), because the context of the implementation of a program seemed to affect the result in a substantial way. As explained later, we need not be so pessimistic in general (i.e., for other fields) or about whether methods can be identified and used to deal with a reformulated construct notion and with the uncertainty of inferences from ODs. Full coverage and perfect generalization may be a far ideal, but imperfect generalization should not prevent us from making useful inferences, albeit associated with more uncertainties than if constructs were unitary and could be represented as fixed points in the construct space. In contrast to perfect generalizability, imperfect generalizability implies variability of findings.

Other Disciplines

The treatment of constructs in other disciplines are both similar and different from the C&M approach. In physics, constructs are physical properties referring to quantities defined by equations that specify precise relations among quantities (Mari et al., 2019). $E = mc^2$ is an equation expressing the relation between energy (E), mass (m) and the speed of light (c). The C&M nomological network is a weaker version of relations among constructs explicated in physics, with weaker quantitative properties and less precision. In contrast with the constructs in physics, Cartwright and Bradburn (2011) and Cartwright and Runhardt (2014) used the German term “Ballung” to characterize social science concepts as agglomerative and multifaceted, which we term compositeness, and as context dependent, which we consider a form of variability of content. Content validity, as defined by C&M implies an aspect of compositeness. Content validity is the coverage a test provides of a universe of associated items, implying a domain-like nature of constructs. Content validity implies that items are not simply

exchangeable. Koch (1992) used the terms “open horizon” for constructs and “practical infinity” for the universe of possible ODs in psychology, which implies that the full content cannot be practicably captured. Although a clear definition of construct is important, it does not prevent the fact that in reality, a construct is composite, complex, and not fully captured by a measure (VanderWeele, 2022). While the definition of constructs specifies one or more defining properties, other, secondary properties of instantiations imply that measures do not cover the whole underlying complexity. C&M’s phrase “assumed to be reflected in test performance” does not specify how well the construct is reflected, which fits with the idea of the approximate nature of ODs.

Although the C&M definition of constructs is vague, its methodological approach to construct validation through psychometrics (correlations, factor analysis methods, group differences, etc.) has dominated psychological research for 70 years with little debate, even though sporadic discussions have appeared. For example, Smith (2005) reflected on C&M in a special issue of *Psychological Assessment*, but apart from psychometric methods becoming more elaborated and advanced, no substantial changes about the notion of constructs seem to have taken place. Aside from technical psychometric developments in the construct validity approaches, the construct-OD gap remains a potential source of dissatisfaction. The inherent abstractness of constructs (dissociated from specific instances) might lead to the impression that constructs are vague, and the multitude of measures and treatment implementations might lead to the impression that the ODs are not sufficiently crafted and need improvements. This construct-OD gap and resulting frustrations might have motivated the recent trends described above and elaborated here. Common to downscaling constructs and upscaling ODs is that both trends aim to narrow the construct-OD gap, although in seemingly opposite ways.

Recent Trends

In the trend of downscaling constructs by moving away from traditional constructs, *local constructs* are chosen that coincide with the ODs or the data themselves. This trend moves away from viewing constructs as inherently abstract and approximate (Subsections “Specific effects and Model Parameters,” “A Network Approach,” and “Machine Learning and Biological Approaches” below). As will be illustrated, the downscaling of constructs is noticeable in various domains such as social and cognitive psychology, mental health, and in various approaches such as machine learning, biological approaches, and network models (to be differentiated from the C&M nomological network). Although this trend does not (yet) dominate psychological science, it might be gaining momentum and could ultimately threaten the role of constructs in the field as explanatory and integrative notions. By *downscaling constructs* to the level of effects and observables functioning as local constructs, this trend engages in a fallacy that naming an effect explains the finding or it suggests that explanation and

integration are not important. Below, we illustrate this trend with examples.

At first glance, the contrasting second trend toward *upscaling (improving) ODs* to better approach constructs seems antithetical to local constructs. Upscaling ODs, in contrast to downscaling constructs, includes describing constructs more precisely so that ODs can be better determined for measuring these constructs (Subsection “Calls for Renewed Attention to Definitions and the Quality of ODs in Measurement”). Proponents of *upscaling ODs* instead of downscaling constructs expect better research quality and replicability. Although upscaling ODs seems counter to downscaling constructs (i.e., local constructs), these opposing trends might be similar in not easily allowing for the accommodation of the complexity and variability of psychological phenomena. Both downscaled constructs and constructs measured by upscaled ODs might hamper the explanatory and integrative role of constructs as abstractions across heterogeneous sets of findings.

There is another way to address the problems that spawned upscaling ODs and downscaling constructs that preserves earlier notions of constructs as domains from which to sample, that remains consistent with the C&M view. As outlined by C&M, earlier notions also considered constructs ranging from close to data (e.g., the Zeigarnik effect as a local construct for better recall of incomplete tasks) to abstract theoretical (e.g., ego involvement measured through recall of incomplete tasks, C&M, p. 289). In Subsection “Critical Discussion,” we examine these issues more closely.

Specific Effects and Model Parameters

In the past 20 years, there is a general trend of less theorizing in psychology (Cummins, 2010, McPhetres et al., 2021) even though theory development and testing remain the ostensible purposes of research. The longstanding problem is described by Clarke (1987), who refers to “pieces of a jigsaw which accumulate in journals” (p. 35), without an eye for the picture the pieces depict. Instead, the field has shifted focus to examining specific effects and relations, as amply illustrated by literature on the so-called replication crisis (e.g., 13 different effects in Klein et al., 2014; 97 effects in Open Science Collaboration, 2015). The effects are from the domains of social and cognitive psychology, and the focus is exclusively on specific effects, not on underlying theories and constructs. The concern for replication is accompanied by large-scale direct (rather than conceptual) replication studies of specific effects (McShane et al., 2019), as promoted recently by several journals including the *Journal of Personality and Social Psychology* and *Advances in Methods and Practices in Psychological Science*.

Affording primacy to specificity (in design, variables, model specification, predictive patterns, markers, etc.) is illustrated by calls to make direct replications a priority (e.g., Flake, 2021; Simons, 2014; Zwaan et al., 2018) compared with conceptual replications (e.g., Crandall & Sherman, 2016; Petty, 2018). With direct replications,

constructs move to the background and effects, parameter estimates, and predictive patterns move to the foreground. The criticism that effects are nonreplicable typically emphasizes research design, data collection, statistical analyses, and reporting of empirical findings. In contrast, it is less typical to observe criticism of the constructs and theories in which effects are embedded.

A focus on specific effects and relations among variables appears in multiple areas of psychological science. Two examples from social psychology are (a) the “mega-studies” approach (Milkman et al., 2021) to evaluate specific operational manipulations and their effectiveness for persuasion rather than conceptual bases of the treatments, and (b) a downscaling of the implicit attitude construct to attitude assessments with indirect measures (Greenwald & Lai, 2020, p. 439; for context see Petty & Cacioppo, 1981) and to specific findings obtained with those measures.

An illustration from a different domain is that cognitive models can become increasingly specific depending on the experimental context and might thus lose their generality. Parameters from these context-specific models are assigned the explanatory role of traditional constructs (e.g., Evans et al., 2019; Lee et al., 2019). Consider Evans et al. (2019), who demonstrated that supplementing choice data with observed response time data enhances the choice data and improves parameter recovery and prediction of effects.

Similarly, computational approaches in psychiatry and neuroscience, which depend heavily on machine learning and other bottom-up approaches to deal with high dimensionality and complexity, focus on specific model parameters and specific predictive patterns to identify mental function and dysfunction (e.g., Huys et al., 2016). A similar example can be found in the study of ideologies, such as predicting conservatism and dogmatism based on specific parameters from cognitive task models (e.g., Zmigrod et al., 2021). Also, psychiatric syndromes and ideology are sometimes downscaled to specific parameters in a computational model.

Focusing on specific observed effects and estimated parameters to model specific tasks can both forgo the added value that constructs provide for interpreting effects and parameter estimates. In this context, constructs can presumably be done away with because there are no apparent repercussions for constructs (or theories), nor do constructs add meaning to the specifics of the study.

A Network Approach

The network approach (Borsboom, 2017; Borsboom et al., 2021), which is presented as an alternative to constructs and latent variables is also a trend toward the specific (downscaling constructs)—closer to observable variables (Schmittmann et al., 2013).¹ Nodes in the network are no longer conceptualized as abstract latent constructs (cf. C&M’s approach). Instead, they are items, symptoms, specific behaviors,

¹Although network models are often presented as different than and advantageous compared to latent variable models, it has also been suggested that the two approaches have much in common and reduce to each other in at least some circumstances (see Burns et al., 2022).

opinions, or feelings. Previously, such specific nodes would serve as measured indicators of constructs; in this network approach, however, these specific nodes are assigned greater primacy as important in and of themselves. Constructs have no further role than umbrella labels (a descriptive label whose role is to name a set of heterogeneous observables). Any explanatory role for constructs is preserved for a new type of node but in a different kind of network than that of C&M. This new kind of network is now made up of specific variables that are much more concrete than the construct of which these specific variables used to be indicators. For example, symptoms of depression and their causal relations replace depression as a construct (Borsboom, 2017; Boschloo et al., 2016) and specific feelings and beliefs and their causal relations replace the attitude construct (Chambon et al., 2022; Dalege et al., 2016). Unlike C&M's nomological net, the network here is at the level of specific variables—not constructs in sense of C&M. A result of downscaling constructs from a network perspective is less room between observations and network nodes, shifting away from tolerance of uncertainty in C&M's construct definition.

Machine Learning and Biological Approaches

Biological approaches and machine learning also depart from the traditional role of psychological constructs. These approaches are not explicitly critical of traditional formulations of constructs but are considered promising avenues with their own merits.

Machine learning is a collection of bottom-up methods that take a brute force, predictive data-analytics approach, with observables modeled directly as inputs and outcomes. The black box is considered sufficient to replace mechanisms that generate phenomena. Accuracy of the prediction—not theory—is given primacy. For example, markers (whether biological or behavioral) of disease need not be accounted for by any explanatory mechanism connecting them to outcomes (illness). The lack of an explanatory mechanism is mainly pragmatic and not based on epistemological skepticism of constructs. In general, predictions in machine learning do not require constructs, which are sometimes seen as nuisances.

Biological approaches might also replace constructs with observable genetic, neural, hormonal, and/or physiological functions. In some cases, these observables replace constructs as explanatory variables (e.g., IL-6 and its role in inflammation and depression). Alternatively, one can view constructs as complex weaves of biological variables, psychological variables, and behavior, in which the specific biological indicators function as explanatory variables. Explanations need not be one-directional (i.e., from biology to behavior), given the many well-documented transactions between biology and behavior. For example, depression can be a product of inflammation, and depression worsens inflammation (Miller & Raison, 2016). Other research shows bidirectional effects of parenting on male externalizing (antisocial) behavior of children (e.g., Beauchaine et al., 2005). At the same time, when children who are at high genetic

risk for antisocial behavior are adopted away at birth, they might elicit negative parenting behaviors in their adoptive parents that reinforce their antisociality (O'Connor et al., 1998). We note, however, that integrative theories that specify constructs across biological, emotional, and psychological levels of analysis exist, can broaden our understanding of mechanisms, and serve to organize observable data across levels of analysis (e.g., Beauchaine et al., 2017).

Calls for Renewed Attention to Definitions and the Quality of ODs in Measurement

Two sources considered as inducing non-replicability of findings are unreliability and invalidity of measures (e.g., Asendorpf et al., 2013; Fabrigar & Wegener, 2016; Flake et al., 2017; Funder et al., 2014). Machery (2021) formulates the problem more generally as a mistaken confidence in data, including the special case of a mistaken confidence in measures. It follows that measures should be chosen with care and the process of this choice is made more transparent (Flake, 2021; Flake & Fried, 2020; Fried & Flake, 2018; Lilienfeld & Strother, 2020). We cannot agree more.

Problems raised by the articles cited above can be interpreted as calling for upscaling measurement-based ODs. These articles raise three important points: (i) the jingle fallacy, (ii) unreliability, and (iii) invalidity. Both Flake and Fried (2020) and Lilienfeld and Strother (2020) noted that the same name of a measure does not guarantee it captures the same construct (the jingle fallacy). Flake and Fried (2020) and Lilienfeld and Strother (2020) also express concerns about unreliable measures, which are often used in psychological research (Barry et al., 2014; Flake et al., 2017). Invalidity might be the most troubling of all. Invalid measures do not and cannot answer the questions we seek to understand. Flake and Fried (2020) offer a list of validity-related questions for researchers to answer when measuring psychological constructs. The authors “do not dictate the best or correct way to create or use a measure, how to evaluate the validity of a measure, or which validity theories or psychometric models to use” (p. 459), but they rightly insist that measurement instruments be closely in line with the construct (“instruments that measure it [the construct] accurately,” p. 456). In sum, crucial points in upscaling ODs are the definition of the construct and matching a measure with the construct.

Critical Discussion

Narrowing the construct-OD gap by focusing on specific variables and effects, as one can notice in the replication literature and other literature on effects, in network approaches, and in machine learning and biological approaches, can lead to new and fruitful conceptualizations. Unfortunately, narrowing the construct-OD gap might also lead to fragmentation. Local constructs might become too narrow for generalization and integration within broader theories. We also question whether being more specific can

resolve the replication crisis because specificity does not guarantee that we measure the same variables or effects across studies, as illustrated by violations of measurement invariance. Psychological variables cannot be perfectly pinned down with sharper ODs or more well-defined constructs. To better reflect and understand real-world phenomena, we need more flexibility in our approach instead.

We do not object to calls for better measurement; what we question is how to consider the phrase “accurately measure it [the construct].” Conceptually, this seems to imply hitting a point target such as a bullseye (although not logically implied, it is an appealing interpretation). The situation is different if constructs are composite abstractions and cannot be reduced to a point. An accurate measure would then require that the observables cover the unobserved construct more broadly. We highlight this point-target analogy between observables and constructs because this conception is inconsistent with the notion of content validity, which implies some degree of compositeness that is better represented as an area than as a point.

The concept of content validation has nearly disappeared from the literature, possibly because it is inconsistent with the implicit ideal of constructs as perfectly homogeneous. Still, content validity remains an important type of validity in *Standards for Educational and Psychological Testing* (AERA et al., 2014). C&M were very explicit in their language describing tests as a sample of items from a universe. In their view, whether items sufficiently represent the universe is established deductively (p. 382) based on the universe definition. As explained, this has been further elaborated in Cronbach’s generalizability work (Cronbach, 1982; Cronbach et al., 1963). The content validity concept is meaningless if items are *exchangeable* indicators of a latent variable (i.e., exchangeable indicators simply hit the same point on the target) ignoring the independent residual (e.g., a unique factor, which does not count for measurement). The only difference between exchangeable items is the strength of their link with the factor (factor loading). Assuming exchangeability, weaker links can be compensated by adding items. The corresponding latent variable as the approximation of a construct always is the same point (a vector in a geometric space—this is not a metaphor but a latent variable representation) independent of the set of items. Without exchangeability, a different set of items would define a different point when combined and together they would define an area. If different sets of items and their corresponding latent variables are like bullets shot at a target, we consider content validity would be the area that is covered by samples of items and the corresponding latent variables. Without perfect exchangeability, content validity of tests must be conceptualized as an area that is covered and must be sampled in order to measure a construct. We call the area the content of the construct.

Consider different measures of depression, which are correlated positively but only moderately. Do these measures then represent different constructs? Or do they partially cover the same construct (cf. the compositeness nature of constructs)? Definitions of constructs can be misleading

when their extension is ignored: “Intensional [as opposite to extensional] characterization of a domain is hazardous since it selects (abstract) properties and implies that new tests sharing those properties will behave as do the known tests in the cluster, and that tests not sharing them will not” (C&M, p. 292). Definitions of constructs are intensional characterizations and refer to defining properties of the construct. If a construct is defined in terms of properties, the extensional content (i.e., the actual items that could be described as representing the construct) might share the definitional properties but sharing the definitional properties does not make the content elements exchangeable because they might be heterogeneous in terms of other, secondary properties.

An important argument for upscaling ODs (better measurement) is that in practice the definition and measurement of constructs can be afflicted with questionable practices and nontransparency (Flake & Fried, 2020). We agree that questionable practices should be countered and transparency is necessary. However, some questionable practices (e.g., ad hoc dropping of items that do not load strongly on a factor) might stem from imposing unrealistic ideals for measuring constructs. Other “questionable” practices, such as using different measures for the same construct in replication studies or a meta-analysis, are not necessarily questionable from a generalization perspective (assuming that items are not exchangeable). What is called the jingle fallacy is not always a fallacy. It can be a proper reflection of the same construct but partially covered by different items in different ways. In the same way, if we consider experimental manipulations and instantiations of the independent variable as sampled from a broader set (i.e., as items of a test are), they are also not necessarily perfectly exchangeable. In principle, one can define a set of implementations (a domain) of the independent variable to which a hypothesis applies (Simons, Shoda, & Lindsay, 2017). Reflecting the set of implementations is the equivalent of sampling item sets from the universe associated with the construct.

Both trends (downscaling constructs, upscaling ODs) imply that the associated constructs do not suffice for more than an explanation of specific psychological phenomena in a specific type of study. A reformulation of constructs as areas is broader and more flexible: the construct-OD gap is not a problem to resolve but treated as an essential and even useful feature for generalization instead of parceling out psychological research into disjointed effects and specific findings. In a later section, we present a reformulated notion of constructs. First, however, we discuss the history and epistemology of constructs (Section “History and Epistemology of the Construct Notion”) because it elucidates how post-behaviorism psychology has evolved. Post-behaviorism began by identifying concepts with their operational definitions and (relations between) observables and later evolved to include constructs that are unobservable, hypothetical explanatory notions (hence the construct-OD gap) with an epistemological status that can differ depending on the view of researchers. We also characterize beneficial roles constructs with different breadths can play (Section “The

Role of Constructs”) to provide a frame used to reformulate the notion of constructs.

History and Epistemology of the Construct Notion

History

Psychology’s move away from early behaviorism to acceptance of the C&M construct approach occurred gradually. Behaviorism started out with the assumption that only observables were amenable to scientific study (see Skinner, 1963) but constructs were eventually embraced by behaviorists (e.g., Patterson, 1986). An early step came from a *Psychological Review* article by Stanley Smith Stevens, (1935), titled “The operational definition of psychological concepts,” in which he distanced himself from behaviorism and reintroduced *concepts* (the term “construct” was not yet used). Stevens equated concepts with observables (i.e., there was no construct-OD gap). To Stevens, the meaning of a concept is what the concept denotes: “the criteria of its applicability or truth consist of concrete operations which can be performed” (p. 517). This point of view was almost literally the same as Bridgman’s (1927) formulation in “*The Logic of Modern Physics*”: “we mean by any concept nothing more than a set of operations; the concept is synonymous with the corresponding set of operations” (p. 36). For Bridgman, an operationalism view on concepts was the result of an analysis of the progress made in physics, whereas for Stevens it was a program aimed at setting empirical psychology on the rails. On the one hand, Stevens wanted to use concepts, and on the other he indicated that criteria for scientifically based conclusions should stem from observations. Taken together, and motivated by avoiding subjectivity, Stevens wanted to use concepts such as those used before behaviorism while ensuring objectivity of the discipline because observability is crucial for intersubjective agreement.

Around that time, inspired by behaviorism, neo-behaviorists such as Tolman (1932, 1938) introduced the notion of “intervening variables,”—functions that connect observables with other observables, where observables referred to stimuli and responses. Although “intervening variable” nowadays suggests a mediator, this was not Tolman’s original meaning. Rather, he was referring to the mappings of observable variables with one another *in the absence of an additional intermediate variable*. Thus, when mapping X on Y , the possibly varying relation between X and Y is itself the intervening variable (there is no third, mediating variable M).

One interpretation of an intervening variable from our current perspective is a random regression coefficient (i.e., a random slope)—regressing Y on X across repeated measures within an individual or across individuals within a group (in a multilevel model). This coefficient varies as in a mixed model, and therefore is itself a variable describing the relation. Hull (1943) later expanded Tolman’s idea. For example, a concept such as the approach gradient is an intervening variable defined by the relation between the distance of an organism to a food target and the locomotive force of the organism approaching the food (the smaller the

distance, the stronger the force). Depending on the organism, the food, and the situation, approach gradients vary.

A related intervening variable is ambivalence, defined by a specific kind of relation for approach and avoidance behavior toward a target object: the approach first strengthens if distance is reduced but then turns into avoidance when getting very close to the target, followed by a new approach when the distance increases. The repeated approach-avoidance-approach pattern is induced if the organism has experienced both reward (satiation) and punishment (electric shock) in earlier encounters with a food. As an intervening variable, ambivalence is the variable relation between distance and approach.

Gradually, intervening variables were also viewed as concepts to explain relations between variables. For example, ambivalence as a concept can be invoked as a source of switches between approach and avoidance as a function of distance to food. When ambivalence acquires surplus meaning as a concept for hypothesized underlying processes (not reducible to just a relation between distance and approach), it simultaneously becomes a hypothetical concept *and* is more difficult to operationally define (harder to define than stimulus-response relations such as approach gradients). Over time, researchers started to introduce concepts defined in terms of unobservable stimulus-response relationships, such as attitudes being defined as an implicit (unobservable) variable connecting responses to stimuli (Doob, 1947).² In Lewin’s (1942, 1943) Field Theory, observed variable relations were theorized to result from forces in the psychological field (also called life space). These (unobservable) forces and their interactions explain (observable) responses to external stimuli.

This shift from intervening variables as observable relations to hypothesized sources of relations between variables led MacCorquodale and Meehl (1948) to introduce a new term—“hypothetical constructs”. Such constructs carry surplus explanatory meaning, whereas intervening variables are not more than observable relations between observables. Intervening variables describe and do not explain. MacCorquodale and Meehl insisted that the term “intervening variable” be preserved for functions mapping observables on other observables, without an explanatory role, whereas “hypothetical constructs” should be restricted to concepts only partially captured by observables with a “hypothetical” explanatory role. “Surplus meaning” of “hypothetical constructs” is the crucial difference from “intervening variables” (p. 106).

Introduction of “hypothetical constructs” in empirical research gave rise to the foundational C&M (1955) paper on construct validity but omitting the qualification “hypothetical.” Importantly, the idea that constructs extend beyond their observables leaves room for the explanatory value of constructs vis-à-vis their indicators (i.e., “A construct is some postulated attribute of people assumed to be reflected in test performance”, C&M, p.283), and vis-à-vis

²Note that for Doob, attitudes were the intervening variable. We are not referring here to the implicit attitude test.

other constructs in the structure. It is important to note that the qualifier “hypothetical” was replaced by “postulated.”

Slaney and Garcia (2015) explained that overall, the notion of constructs is ambiguous in psychological science. Slaney (2017) considered different conceptual ways that constructs have been interpreted since C&M: for example, from a set of variables (Nunnally & Bernstein, 1994), to constructions for approaching traits “which exist prior to and independently of the psychologist’s act of measuring” (Loevinger, 1957, p. 642) hoping they are sufficiently valid representations of traits; to entities with causal power (Crocker & Algina, 1986). Except when constructs are assigned a purely descriptive and denotative role, there is an inherent gap between the construct and the OD, and as far as constructs are explanatory, they are necessarily hypothetical. In Section “The Role of Constructs,” we will explicate the roles we expect constructs to play.

The Epistemology of Constructs

Borsboom (2005) distinguished between two general beliefs about constructs: realism and anti-realism. To illustrate these positions, we use an attitude as an example. Attitudes are evaluations that fall along a valence continuum (i.e., from extremely unfavorable or negative to extremely favorable or positive). *Realism* implies that constructs exist (they are entities) and have causal effects on observables. In the attitude context, this means that certain behaviors follow from an existing attitude that is separable from the feelings and behaviors it causes. For example, the idea is that a person could hold a positive attitude toward chocolate ice cream, and that positive attitude would make the person more likely to choose chocolate ice cream over other options there are at an ice cream store. Such separation implies that, in principle, the attitude can be established without relying on the manifested behaviors it is meant to predict. As a cause, it does not (and should not) coincide with its effects.

Following Borsboom (2005), there are three important *anti-realist* viewpoints reflected in philosophy of science. *Logical positivism* views constructs as logical constructions with correspondence rules. For example, it is possible to formulate a statistical model for covariation across individuals (or situations) for a set of responses that are commonly interpreted as reflecting the attitude. The set of rules then sets the bounds of the construct in and of itself. Logical positivism does not imply preexistence or causality. According to *instrumentalism*, constructs are instruments with predictive value when they are measured in some way. In this view, attitudes need not be causal, and need not be embedded in a model with rules for covariation between indicators. Predictive utility is the only requirement. For example, a positive score on a measure of attitude toward a candidate for office might be predictive of donations and voting behavior. In instrumentalism, the predictive relation itself defines the construct. For a realist, however, a positive attitude toward the candidate would be considered as the

cause of these behaviors. Finally, in *social constructivism*, constructs are formulations in our minds and in the writings of theorists and investigators. Attitudes are constructed to make sense of others’ behaviors, which could differ depending on ascribed meaning based on interpretation of the context. For example, seeking detailed information about a candidate for political office could be motivated by positive or negative attitudes. Given that the same behavior can have different meanings and that different behaviors can have the same meaning (e.g., attending a meeting because one likes a candidate or not attending a meeting because one is sufficiently informed about a candidate one likes), social constructivism and its interpretative flexibility are often appealing. To reach this level of flexibility with common measurement approaches, an infinite number of moderator variables might be needed.

Readers will note that these views correspond to different historical eras in philosophy of science. The C&M view on constructs “as expressed in test performance” is a step away from logical positivism toward realism. The touch of realism in C&M is unsurprising; C&M cite Feigl’s (1950) ideas on logical empiricism, which is an intermediary between logical positivism and scientific realism as explained by Slaney (2017, Chapter 6). Next, we present a working definition of constructs with desiderata for the roles they play, without strict adherence to a single epistemological view. However, instrumentalism can hardly be reconciled with the desiderata. The roles of constructs can be considered as marginal conditions for our reformulation of the notion of constructs in Section “Construct-Related Issues and Possible Solutions.”

The Role of Constructs

We begin with a working definition of constructs with four features. Having four features is rather elaborate and contrasts with straightforward definitions, which likely follow from the notion that constructs are basic or primitive. The adjective “primitive” refers to a notion from philosophy and formal systems as intuitive ideas—not to their structure vis-à-vis other concepts. An elaborate working definition is not only more comprehensive, but also sheds light on various functions that constructs can take on. Our working definition is illustrated with three brief construct presentations in Section “Three Constructs: Attitudes, Autism, Evidence Accumulation (Drift Rate)”: the social psychological construct of attitudes, the social and clinical construct of autism, and evidence accumulation (the “drift rate”) in the drift diffusion model. Below, we provide a brief description of each example construct and then delve into the four features of constructs illustrated by these example constructs.

Three Constructs: Attitudes, Autism, Evidence Accumulation (Drift Rate)

Attitudes

Attitudes are generally defined as overall evaluations of objects (e.g., Eagly & Chaiken, 1993; Petty & Wegener,

1998). Evaluations of objects fall along a valence continuum and are generally measured with responses to items asking people the extent to which they view the object negatively or positively. Such items can take on various forms, and the methods used to assess the items can vary across scale types (for a review, see Wegener & Fabrigar, 2004). Traditionally, attitudes have been treated as hypothetical (latent) constructs that influence responses to items in evaluation scales for direct measures of attitudes. Alternatively, indirect (sometimes called implicit) measures assess behaviors or judgments that do not directly ask for evaluations but that are presumably influenced by such evaluations. Typical examples range from older projective tests (in which people's descriptions of ambiguous stimuli are thought to reflect underlying evaluations) (e.g., Proshansky, 1943) to the more contemporary Implicit Association Test (in which consistency of evaluative associations between two categories of stimuli speed up responses in a categorization task) (Greenwald et al., 1998); see Fazio et al. (1995) for a conceptually similar evaluative priming measure. To some degree, both direct and indirect (implicit) measures of attitudes have taken a factor-analytic approach in which the latent construct of attitude is assumed to influence responses that can be taken together as an indication of the favorability of the underlying attitude.

Attitudes have played many roles in psychological studies and theories. Attitudes can be induced by affective and cognitive processes and be composed of affective and cognitive content, which serves as the basis of the attitude toward an object (Crites et al., 1994). The composition of an attitude might differ depending on the object and other factors. Attitudes serve as the primary outcome (dependent measures) of studies examining how attitudes are formed or changed (see Petty & Wegener, 1998, for a review). In other cases, attitudes are the focal independent variables used to predict and explain related thinking or behavior (e.g., Wallace et al., 2020). Attitudes serve as key explanatory variables in connecting interventions such as marketing campaigns, political campaigns, or health communications to behavioral outcomes. The examination of attitudes in psychological theory and applied research is likely due to their potential explanatory value and potential causal effects on behavior.

Autism

An additional way in which constructs are applied is in the case of clinical presentations, such as mental health or developmental conditions. One illustrative example is autism. Autism is a construct defined by a group of impairments in social communication and repetitive behavior (American Psychiatric Association, 2013). In addition to being defined by these core features, diagnosis of autism has both predictive and explanatory power for a number of behavioral indicators (e.g., repetitive motor behaviors, language delay, challenges in daily living).

Given that an autism diagnosis is commonly used for practical purposes, such as access to disability supports, special education services, or inclusion in research studies, a

fixed definition is required for practical purposes—despite the wide acknowledgment that autistic features lie on a relatively normal distribution in the population, without an obvious dimensional “cut point” (Landry & Chouinard, 2016; Sucksmith et al., 2011). This lack of a cut point presents a practical challenge and tension between the empirical search for the “true” nature of autism and the need to settle on a universal definition for clinical purposes; perhaps as a result, the construct of autism has undergone repeated definitional modifications over time—which is likely to continue given concerns in the research and clinical communities that current definitions remain imprecise in a way that compromises research and clinical care (Lord et al., 2022; Singer et al., 2023). A more extensive discussion and the practical relevance of the construct is provided in the [Supplementary Material](#).

Evidence Accumulation

A third way in which constructs can be applied to understand observables is through cognitive modeling. Cognitive models are constructed from overarching theories about the mechanisms and processes that underly a decision-making process. The mechanisms and processes are considered constructs within the model. Often, the models are structured such that the construct itself is represented as a parameter (or set of parameters), and these parameters explain different patterns of observables from an experiment. The process of fitting a model to data entails identifying the combination of parameter values that best match the data.

One example of a cognitive model that uses constructs to explain how decisions are made is the Diffusion Decision Model (DDM; Ratcliff, 1978). The DDM is based on an overarching theory called sequential sampling theory, from which many other models are also derived (Brown & Heathcote, 2008; Busemeyer & Townsend, 1993; Krajbich & Rangel, 2011; Merkle & Van Zandt, 2006; Nosofsky & Palmeri, 1997; Usher & McClelland, 2001). The DDM assumes that each choice alternative is represented as a boundary in an evidence accumulation space. At each moment in time, observers collect noisy samples of evidence for each alternative, and these noisy samples are added up over time to provide total evidence for each alternative. The rate at which evidence is integrated is a parameter referred to as the “drift rate,” and this parameter is an important construct within the model. Once enough evidence favoring a particular option has been gathered (e.g., once the total evidence for a response option exceeds a prespecified amount), a decision is made corresponding to the choice of that option. The rate of evidence accumulation explains the response time and response accuracy.

In the next section, we highlight four features of constructs related to the roles they play (i.e., constructs are embedded within structures, explain behaviors, have more than an explanatory function, and are theoretical and practical). We also describe how these features are evident in the three constructs of attitudes, autism, and evidence accumulation.

Constructs Are Abstractions Embedded in Componential Structures

A construct is an abstraction, a composite in a structure that usually consists of more elementary and less abstract components. In a very broad sense, the structure is a theory. The theory with its components and composite can be expressed verbally without a formalization or can be formalized in a statistical or mathematical model. Unlike in physics, psychological theories are usually not expressed in mathematical terms. Attitude and autism theories are typically verbally (and not mathematically) expressed. By contrast, the DDM for binary decisions is a formalized theory with a mathematical expression.

Having a componential structure is an essential and desirable feature of constructs: constructs inherently lose their meaning when stripped of related components (i.e., within-construct components [e.g., subconstructs as represented by multidimensionality] and other constructs in the structure). Relations within the structure constitute necessary integration of a theory. Unlike in physics, the relations about psychological constructs generally cannot be described in precise terms (but they are in cognitive models such as the DDM). The imprecise nature of constructs could be one reason for a trend toward specific variables and effects in the hope that precision will improve.

Constructs Function to Explain Behaviors

Explanation is to give a sense of meaning to something observed by referring to something (not directly) observable or unobservable and more abstract (i.e., the construct) that is part of a broader system (i.e., the structure). This broader system could be a model, a metaphor, a set of related concepts, a pattern, or a narrative, etc. Explanation is not necessarily causal. In a very broad sense, explanations are all theories. Whether an explanation is causal or not depends on one's epistemological perspective, for example a metaphor or a narrative is explanatory but not necessarily causal. Furthermore, a metaphor or narrative need not be literally true (as in realism).

Referring to the three example constructs, the explanation provided by constructs can be of different kinds: causal (attitudes are often considered as causes of related behaviors), patternlike (deficits in social communication are part of an autism pattern), or a parameter within a statistical model (drift rate is a parameter in a model for response time and response accuracy). Constructs can have explanatory value for observations as well as for other constructs, as illustrated by the example constructs. Although constructs can have explanatory value for other constructs, the set of constructs involved also need an explanatory role for observations. Without such a role, any structure of constructs is meaningless. When constructs are implemented in factor (psychometric) models, the common factor model is reflective in that factors are assumed to explain variance in their observable indicators. The explanatory role can be indirect, as in a higher-order factor model with a general factor that explains relations among lower-order factors. Alternatively, the factor

model might be formative instead of reflective, so that the factor is instead defined (formed) by its specific set of indicators. In the case of formative models, the factor needs to explain external observables directly or indirectly through another factor for the formative factor to be identified (Blalock, 1971; Bollen, 1989; Land, 1970). For a theory to have an explanatory and integrative role, some of the constructs involved are likely to be broad (global instead of local). Attitudes and autism are clearly broad constructs and evidence accumulation is as well if not narrowed to the drift rate parameter for a specific decision task.

Constructs Can Serve More Than an Explanatory Function

The explanatory nature of constructs does not preclude other functions of constructs. For example, constructs can also predict. Prediction can be related to explanation, but not all predictive relations are explanatory. Thus, constructs do not just have predictive functions alone. For example, cognitive ability and personality have explanatory *and* predictive value. The same is true for attitudes and autism whereas drift rate has primarily a model-based explanatory value. Based on the crucial explanatory role of constructs, something with predictive value but no meaningful explanatory function does not qualify as a construct. For example, some bio-markers have predictive value for suicidal behavior without any explanation for why they are predictive (e.g., Sudol & Mann, 2017). Although such bio-markers might be very useful for preventive purposes, they are not constructs in that they cannot explain why suicide is attempted.

Because constructs are abstractions (Tal, 2020), they can also have a descriptive role, as a bundle of behaviors with one or more common properties. For example, the act-frequency theory states that traits are sets of acts (Buss & Craik, 1983). If that would mean that traits are nothing more than descriptive categories of acts, in line with the Nunnally and Bernstein's (1994) construct notion, it would not suffice for traits to be constructs, even though they might have predictive value. Abstraction by itself does not imply explanation, but abstractness allows for constructs to play an optimal explanatory role by being integrative. An explanation (stemming from a construct) cannot change too much across individual instantiations or data sets. For example, the attitude construct is rather abstract and applicable to a broad range of studies with different designs and implementations without required changes in the defining properties of the construct. The abstraction inherent in a construct allows for the construct to be applied across different modes: time, situations, behaviors, etc. Constructs require some minimum level of abstraction, and this abstraction should be flexible enough to allow for explanation in the presence of variability across different modes.

The Role of Constructs Can Be Theoretical and Practical

Constructs are explanatory notions and are used primarily in theories. For instance, attitudes explain why humans

behave and make choices differently. The construct of attitude is a part of broader theories on opinion formation and is itself linked to the culture in which people were raised. Additionally, even though constructs are explanatory, they can be used for practical purposes related to their real-world predictive and organizational functions. For example, attitudes are involved in various types of campaigns (in marketing, politics, health), and the construct of autism can be used to make practical predictions involving decisions on access to treatment, and to investigate treatment efficacy.

Theoretical and practical purposes are often combined in scientific inquiry. Psychiatric syndromes can function as practical constructs for diagnosis and treatment, and types of impairments (as part of the construct) can function as criteria for financial subsidies for education and to support living. At the same time, these constructs might function as components in broader scientific theories. They receive meaning from being embedded in a sensible structure. Being embedded in a broader structure implies that constructs also have an explanatory function because they are linked to other components in the structure and can help to explain these links. We reiterate that the primary feature of a construct is its explanatory function—a minimum requirement of theory. This explanatory function can be theoretical and practical. A practical rule can be a guiding principle, but it is easier to follow a principle if it is comprehensible and relies on constructs that explain why it makes sense to be followed.

Construct-Related Issues and Possible Solutions

For constructs to fulfill their explanatory and practical roles as described in the previous section and to reflect psychological phenomena in a realistic and integrative way, we propose a reformulation of the notion of constructs. By realistic we mean that the complexity and variable nature of the phenomena and the imperfect convergence of findings cannot be explained away as sample variation (sampling error). The reformulation makes the notion more general because it leaves room for a larger variety of constructs and for more flexibility in how they are conceptualized. Narrow and fixed notions of constructs are special cases of a more general category of constructs. The general category includes broader and more variable constructs that can deal with the variable and complex nature of psychological phenomena. We will focus on three extensions concerning the notion of constructs. For each of these we refer to conceptual arguments and empirical findings.

Constructs Are Areas

By areas we mean that the content of constructs is composite and not perfectly homogeneous (i.e., not unitary, not a single point in the bullet-target analogy). We prefer the term “area” to allow for continuous content differences and not only discrete content elements. To draw an analogy with dimensional types of modeling, the geometric space is continuous. Regarding constructs, the notion of their

homogeneity versus heterogeneity can be ambiguous. It is possible that the content of a construct is homogeneous in terms of primary properties (i.e., the properties that define the construct), whereas simultaneously the content is not homogenous with respect to secondary properties.

To map this idea onto psychological constructs, we refer to work by Rips and Conrad (1989) on a taxonomy of mental activities. Participants in their study considered reasoning, interpreting, analyzing as kinds of thinking. Furthermore, thinking was considered a property of reasoning, interpreting, and analyzing. A similar compositeness as for kinds of thinking applies to many scientific concepts in psychology. For example, Gross and Canteras (2012) found that there are kinds of fear: “fear of painful stimuli versus fear of predators and aggressive members of the same species,” (p. 651) and that the kinds of fear depend on distinct neural circuits. We have already mentioned different kinds of attitudes (e.g., affective versus cognitive; Crites et al., 1994). Though an abstract property(ies) might define a construct in the sense that the property applies to all its content, that does not mean that the construct’s content is homogeneous in terms of all properties.

The consequence of heterogeneity for measures of a construct is that the measures mostly cover part of the composite construct. The part these measures cover is commonly heterogeneous as well, in the sense that the measure is constituted by nonequivalent components each referring to part of the construct content. The definition of a measure (e.g., a summed score of a test) fixes the measure operations but it does not guarantee that the components of the summed score are equivalent and neither does the measure guarantee that summed scores have the same meaning for all participants.

Homogeneity in terms of the defining property or properties of a construct might lead to the expectation of perfectly convergent measurement findings. However, pure instantiations of a property do not exist in the sense that instantiations also have other properties. There is no animal that is just a mammal, there is no thinking that is just thinking, and there is no behavior or response that purely and only reflects extraversion. There always is some additional aspect to any instantiation. It might therefore seem as though the construct is not sufficiently precise and that overarching constructs are too vague, suggesting that lower-order and thus narrower (downscaled) constructs are ideal. It is a definitional fallacy to assume that a precise definition in terms of shared properties bestows the construct perfect homogeneity. Perfect homogeneity is not necessary. Homogeneity in terms of shared properties is sufficient to define a construct and can help for integration across the heterogeneity of a construct. The “downside” is imperfect empirical convergence (even after correction for imperfect reliability) unless the construct is completely covered, which would often not be feasible. The “downside” of perfect homogeneity and the ideal of perfect convergence is an infinite proliferation of constructs with a high degree of specificity. The former downside (imperfect empirical convergence) requires a higher tolerance for uncertainty and the

latter downside (perfect homogeneity) requires a higher tolerance for fragmentation.

Bridgman (1927) was already open to the idea that perfect conceptual unity does not apply to concepts in physics, even when convergence between findings with different ODs was reached (Chang, 2021). For psychological constructs, empirical convergence is far from perfect. Compared with psychological constructs, there are not kinds of temperature and length, unless one would refer to higher and lower temperatures and shorter and longer lengths.³ Empirical convergence of these physical measures is not a problem. Temperature and length as properties are separable from other properties. Convergence and the separability of properties in measurement is possible for most physical properties. It is uncertain whether the same ideals of convergence and separability can be reached or should be desired for psychological constructs. Whether these ideals are feasible can be considered a conceptual or a measurement problem, pointing to the need for better concepts and/or better measurement. The ideals and the belief that convergence and separability are feasible serve to explain the motivation for downscaling of constructs (i.e., focusing on local constructs by extending homogeneity to more properties to be more precise) and upscaling of ODs (to capture all the definitional properties; see Section “Recent Trends”). We choose to be open to constructs that are heterogeneous and do not require separability and perfect empirical convergence in our treatment of constructs. Such heterogeneous constructs have the potential to better align with the complex and variable nature of human behavior and to foster integration across a variety of findings.

Figure 1 gives a representation of constructs as areas instead of points. Note that the areas are not sharply delineated if the defining properties do not allow for a sharp delineation. The area representation illustrates two different features of constructs: (a) a form of heterogeneity, mapping onto secondary properties of the construct content (not in terms of its defining properties), and (b) not having perfectly convergent measurements which would present as a scatter of points within a geometrically defined but imperfectly delineated area (cf. bullet-target analogy). The scatter plot does not represent the well-known notion of sampling variability in terms of samples of participants of a study but rather measurement (dis)continuity due to differences in the content of the construct that is represented in the measurement. For instance, developments in estimating parameters in factor models that allow for fuzziness in the solution, which are less strictly confirmatory (e.g., Bayesian approaches with small cross-loadings and small violations of measurement invariance, as presented by Muthén and Asparouhov (2012) and Asparouhov and Muthén (2014), are counter-indications of treating constructs as points in space.

Most constructs in psychology begin as unidimensional entities, but further development tends to evolve them into multidimensional entities. Multidimensionality formally

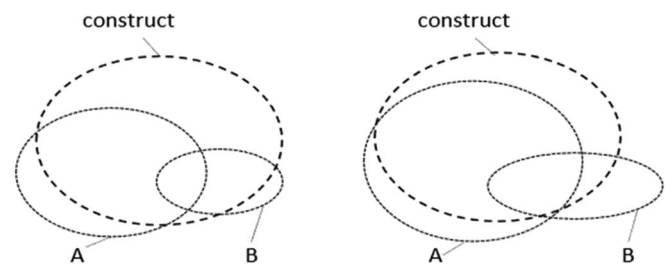


Figure 1. Schematic representation of construct content and measurement continuity. The target construct is represented as two ovals for the same construct, with a dashed line to indicate that its content might not be sharply delineated. On the left and the right, two measures (A and B) are represented with lighter dotted ovals. The A and B ovals indicate parts of the construct to represent partial content that is covered by measures A and B, which illustrates: (1) the hierarchical structure of the construct content and partial possibly overlapping coverage by its measures (A and B), (2) the variable content covered by measures across occasions or studies (i.e., different from left to right) due to variation of the construct content or of each of the measures. Measurement continuity shows in the variation between the measures (A vs B) and in the variation of the same measure across studies, time, etc.

implies heterogeneity of the construct. The Big Five personality traits each contain facets (e.g., Paunonen & Ashton, 2001), which one might treat as more homogeneous subconstructs. However, the items within facets describe even more specific personality nuances that do not map onto error (Mõttus et al., 2017). Furthermore, items might be composite, in turn suggesting an analogy with fractals (Mandelbrot, 1983)—patterns within patterns within patterns. For depression, Fried (2017a) illustrated how seven common measurement scales overlap partially and are comprised of different (partly overlapping) subsets of symptoms. These subsets are not exchangeable, as illustrated by modest correlations (e.g., Bukumiric et al., 2016). For a notion such as evidence accumulation (e.g., drift rate from the diffusion model), it turns out that for different kinds of binary choice tasks, parameter estimates are not highly correlated in the same sample, pointing to low convergent validity even though the estimates refer to the same model and are estimated in the same way (Lerche & Voss, 2017; Lerche et al., 2020). According to VanderWeele (2022) the reality of constructs cannot be captured in one or a few dimensions because constructs refer to a highly complex reality.

We believe that the heterogeneity idea was already present in the content validity notion of C&M. The notion of content validity was introduced by Cureton (1951) in contrast to criterion validity and construct validity, which referred to relations of a test (a measurement instrument) to other variables. However, the focus on relations among variables leaves out the internal content of constructs to be reflected in a measure, as argued by Lissitz and Samuelsen (2007). One can interpret the approximate nature of ODs as emphasized by C&M as one reason for the inherently incomplete coverage of the construct (coverage of a trait in Loevinger’s [1957] perspective).

Constructs Reside in Hierarchies

We propose allowing for construct hierarchies with different levels of specificity as presented in Figure 1 through smaller

³An example of heterogeneity is the concept of force. That there are four kinds of force is a frustration for physicists.

areas roughly within the same construct area. C&M stated that “Constructs may vary in nature, from those very close to ‘pure description’ (...) to highly theoretical constructs ...”. A well-known example of a construct hierarchy is the hierarchical theory of intelligence (e.g., Carroll, 1993). The hierarchical feature of constructs refers to the structure of compositeness of constructs. Multidimensional structures for the same construct are rather common, with each dimension as a subconstruct and specific items at an even lower level (i.e., “very close to pure description” according to C&M). However, Clark and Watson (2019) point out that the hierarchical structure is not a perfectly nested structure in the sense that it can also have overlap. See the overlap in Figure 1, including possible overlap outside the construct. Clark and Watson use the term *interstitial constructs* for blends of constructs, allowing for overlap. The Abridged Five-Dimensional Personality Circumplex (AB5C; Hofstee et al., 1992) is an example of interstitial constructs from the Big-Five personality dimensions literature. Here, neighboring traits in the circumplex overlap. The issue of overlap (and partial coverage) for experimental designs and independent variables as indicators is formulated by Shadish et al. (2002) as follows: “A study’s operations might not incorporate all characteristics of the relevant construct (construct underrepresentation), or they may contain extraneous construct content [overlap with other constructs].” (p. 72). In sum, construct hierarchies do not necessarily have a perfectly nested structure.

In an extension of reliability theory into a multivariate reliability theory, Wittmann (1988) also treats the case of multiple constructs with a hierarchical structure and considers sampling from the set of components (aspects) within each of the constructs involved. Wittmann (1988) forwards that measures capturing more specific aspects of a construct (i.e., smaller areas in Figure 1) are less correlated with other broader variables compared to measures that capture a broader set of aspects of a construct (i.e., larger areas in Figure 1). Empirical evidence for Wittmann’s theory was provided by Kretschmar and Nebe (2021) using the constructs of intelligence and complex problem solving. For example, measures for specific aspects of complex problem-solving correlate less with intelligence than broader measures of complex problem-solving. The controversy about traits being poor predictors of behavior (Mischel, 1968) versus being reasonably good predictors of behavior (Epstein, 1983) is an illustration of the same principle in that measures of specific behaviors refer to less or smaller components of a trait than common personality tests do. How specific a construct needs to be might be determined based on the construct’s research purpose, leaving room for broader constructs to play explanatory and integrative roles, while local constructs play a role for more limited and more specific research purposes.

Constructs Are Variable

The composite content of constructs is not only hierarchical, often it also is variable. To describe how constructs are

variable in content, it might help to recognize that many constructs are macroscopic outcomes based on underlying processes that are relatively more microscopic. For example, different encounters (microscopic processes) with the object of an attitude can lead to different overall evaluations (a macroscopic outcome). Variability in the more microscopic processes can also lead to variability in the content of a construct at the time of measurement (e.g., a more cognitive or more affective attitude) even when the overall level of a construct (e.g., favorability of the attitude) is similar across individuals who experienced different microscopic processes.

To explain the difference between macroscopic and microscopic, we rely on examples from different disciplines. Temperature is a macroscopic property that refers to the average kinetic energy of molecules, whereas the kinetic energy of a molecule is a microscopic property (Mari et al., 2019). The connection between microscopic and macroscopic in the case of temperature is purely quantitative, and does not lead to possible qualitative differences. Compared with the example of temperature, the consequences of microscopic processes for a more macroscopic level of phenomena (the translation from microscopic to macroscopic), is often different for biological systems. For example, how well apples are protected against external pathogenic effects depends on their wax coat, which is naturally produced from within the apple, through biosynthetic processes that are influenced by a complicated set of internal and external factors with possible effects on the composition of the wax. The wax does not only show quantitative effects (thickness and resistance of the wax) but also qualitative differences (composition of the wax with consequences for thickness and resistance).

In a somewhat similar way, the neural circuitry and processes of the brain in interaction with external factors might lead to different kinds of fear (Gross & Canteras, 2012). In Barretts’s (2017a) theory of constructed emotion, the microscopic level refers to interoceptive signals and internal computation in the brain. Depending on the signals and visceromotor predictions (p. 16) and depending on one’s social reality, the same constructed emotion has qualitatively different contents. For attitudes, affective and cognitive processes can each help to determine the attitude’s composition, depending on specific experiences. The network approach hinted at by Fried (2020) with its assumption of dynamic processes of nodes (e.g., symptoms of depression) influencing each other can lead to qualitative differences in the composition of constructs such as depression and their measurements. With macroscopic levels that reflect the outcomes of underlying processes it should then not surprise us that the composition of constructs and their relation with other constructs can vary. The variability of a construct’s content and the varying relations with external variables should not be a reason to abandon macroscopic constructs and measurement.

In fact, there is broad evidence that relations between independent variables and dependent variables do vary and are heterogeneous across psychological studies (e.g., De Boeck & Jeon, 2018; Stanley et al., 2018). This seems true

for correlational as well as experimental studies. Thus, the heterogeneity in results related to the same construct cannot be attributed to a lack of experimental control. A possible explanation for the variation is the absence of measurement (or manipulation) invariance (Fabrigar et al., 2020; Fabrigar & Wegener, 2016). Fried et al. (2016) showed that the same set of depression symptoms can suggest measurement invariance. Yet, violations of measurement invariance are omnipresent (Leitgöb et al., 2023), and factor analysis results often do not replicate (Osborne & Fitzpatrick, 2012). There are several possible explanations, but sampling variation alone is insufficient as an explanation because it has been formally accounted for in statistical testing. Robitzsch and Lüdtke (2022) argue that measurement non-invariance can be construct relevant, referring specifically to non-invariance of the construct itself. The processes underlying a construct are a possible basis for the non-invariance.

To explain sources of non-invariance in a statistical way, Wu and Browne (2015) introduced the notion of *adventitious error*—a form of heterogeneity of the covariance matrix of variables beyond sampling variation that suggests a varying position of psychological variables in the construct space. Adventitious error helps explain why goodness of fit is only approximate and why the measurement of latent variables varies from study to study. As latent variables are model-based representations of constructs, the constructs vary because they map onto a varying covariance matrix. For different sets of indicators and for the same set in a different study, different construct sub-areas are covered, which together define a larger area. Comparing the right part with the left part of Figure 1, one can see change in the areas for measures A and B that represents variability in terms of what the same measures cover of the construct (the same set of indicators may cover changed or different content of the same construct). This is either due to differences in meaning of the same measure or to a variation in construct content. These two possibilities cannot be empirically differentiated.

The above three points of reformulation: (1) compositeness (heterogeneity represented as an area), (2) hierarchical organization of the compositeness, and (3) variability of the compositeness, are all consistent with a variability of measurement outcomes that go beyond sampling variation and is embraced here as a form of *measurement continuity*. Measurement continuity refers to varying measurement outcomes with a common core based on the construct definition. It implies a kind of uncertainty that is not considered problematic but is instead a recognition of the psychological reality. Measurement continuity is a view that differs from attributing variable results and uncertainties to problematic constructs, problematic definitions, and problematic measures. Although it might be difficult to differentiate between suboptimal approaches and the complexities of the object of study (the behavioral reality) as explanations for problems in psychology as a scientific discipline, we believe that more emphasis should be put on the complexity of the psychological reality and on method developments to accommodate the complexity and accompanying uncertainty. We consider

the reformulation of the construct notion as a step in that direction.

What the Reformulation Does and Does Not Do

The reformulation of constructs is not meant as an excuse for suboptimal practices but instead as an encouragement to meet the need for more generalization and integration in psychological research with an openness to more flexible constructs. The criterion of success in this reformulated notion of constructs lies in the results and their contribution to integrative theories with explanatory value. Sharper definitions and more precise measures are desirable if they do not make generalization and integration more difficult.

Our reformulated notion of constructs is consequential and leads to new developments in workable methods. The reformulated notion calls for a generalizability perspective in line with Cronbach's later work (e.g., Cronbach et al., 1963), accepting sources of variation for measurements of the same construct related to sampling from item (i.e., indicator) universes and accepting alternative implementations of the independent variable in experimental studies. From a Bayesian perspective, randomness due to instruments or items is aleatory (e.g., see O'Hagan, 2004) in that increasing sampling (and more empirical information) does not reduce uncertainty. The consequence of a generalizability framework is to embrace uncertainty rather than trying to avoid it or define it out of existence.

Generalization-based methods have been applied in a number of different areas of research design.

- a. For covariance structures such as factor models, Wu and Browne (2015) proposed a multilevel model to formalize the variability of covariance structures. The assumption was that there is a population covariance matrix per study from which the observed matrix is sampled (i.e., sampling variation). The population matrix of a study is itself sampled (call it distorted) from a distribution at a higher level. This approach represents a statistical way to deal with issues of measurement invariance while preserving a coherent structure at a higher level. The higher-order sampling assumes that goodness of fit is only approximate, and the consequence is that standard errors increase under these assumptions.
- b. For the implementation of independent variables in experiments, Baribault et al. (2018) and DeKay et al. (2022) described how meta-studies (a large set of micro-studies) are a sensible and efficient alternative to meta-analysis and can be used as a multilevel approach to generalization based on sampling from the universe of possible ODs for an independent variable. From a philosophy of science perspective, Machery (2020) formulated a similar view in his resampling account of replication. Rather than adding to the direct-versus-conceptual replication controversy (e.g., Crandall & Sherman, 2016; Pashler & Harris, 2012; Petty, 2018; Simons, 2014; Stroebe & Strack, 2014; Zwaan et al.,

2018), Machery proposed to sample from sets of treatments, experimental units (participants), measurements, and settings. Whether a phenomenon replicates can be evaluated only upon resampling, and resampling from treatments, participants, measurements, and settings implies a degree of (aleatory) uncertainty that cannot be reduced by just increasing sample sizes. In our view, the approaches discussed in this paragraph are better in line with conceptual replication approaches than with direct replication approaches.

- c. To determine the power of a study, a generalization approach would mean to consider more than just one fixed effect value. Pek and Park (2019) and Pek et al. (2022) described how to include uncertainty about an effect in power analyses and what various forms of uncertainty might imply about the calculated values. Such approaches lead to more generalized power calculations that incorporate the variability and uncertainty of effects.
- d. To avoid a disconnect or even a competition between macroscopic and microscopic views and approaches in relation to the nature of constructs, multiscale approaches such as those used in systems biology should be considered. In such approaches, microscale dynamics can explain meaningful variations in macroscale phenomena (e.g., Bardini et al., 2017). Fried (2017b) proposed a network approach to psychological constructs as an alternative to reflective and formative latent variables. From a multiscale view, a connection could be made between nodes and processes in the network and latent variables that represent abstract constructs.

Our reformulation and the heterogeneity of construct content do not imply that formative measurement models should be used instead of reflective models. However, MacKenzie (2003) recommended formative models for domain-like constructs (and thus for heterogeneous constructs).

According to MacKenzie, indicator variables define domain-like constructs and changing the indicator variables changes the construct. We believe that domain-like constructs can still be defined in terms of properties (instead of indicators), and property-defined constructs would still leave room for heterogeneity in terms of secondary properties. The defining properties of a construct without perfect homogeneity (i.e., with heterogeneity of secondary properties) can continue to be reflected in the indicators, so that a reflective model can be used. We believe that both types of factor models (formative and reflective) can make sense and that heterogeneity does not determine the choice of measurement model. For example, the covariance matrix for a set of indicator variables can be captured by a reflective model, but that does not imply that the reality of the construct one wants to measure is as simple as the factor model that represents the covariance matrix (VanderWeele, 2022). VanderWeele believes that reflective as well as formative models fall short of representing the complex reality of a

construct and its causal relations. Reflective *and* formative factors change with their indicator variables and, additionally, formative factors also change with their external variables in the model (Edwards, 2011). Rather than to make a choice between reflective and formative models based on the compositeness of the construct, this choice should depend on the kind of relations one wants to model between the indicators, the latent variable, and other variables in a structural equations model. In VanderWeele's (2022) view, the model would most likely be more complex than simply reflective or formative.

Discussion and Conclusion

We have questioned the notion of constructs in several ways while considering recent trends and historical as well as epistemological perspectives. Starting from a working definition of constructs as fundamentally explanatory notions, while also trying to accommodate issues of the construct-OD gap, variable research findings, and related uncertainties, we proposed a reformulation of the notion of constructs. The reformulation concerns (1) a reconceptualization of constructs in terms of areas instead of points in a construct space (allowing for heterogeneity within constructs), (2) a range of constructs from abstract to specific organized in a hierarchical system with possible overlap, and (3) an explicit recognition of variability of the construct content and measures of the construct. The reformulation itself does not eliminate the variability and uncertainty but rather accommodates these issues into a realistically broader notion of constructs that allows for more integration and generalizability-based theory. We believe it is possible to develop methods in accordance with our reformulated notion of constructs.

From a strategic point of view, one might wonder whether a bottom-up approach with specific constructs accompanied by findings that seem more certain, might be a better approach than the proposed reformulation of the construct notion. The advantage of the more bottom-up approach might be that one could rely on what seems to be fixed stepping stones that point the way to arriving at well-founded theory. Only the future can tell. However, the certainty-seeking bottom-up strategy would also require generalization and integration in a later stage. The proposed reformulation of constructs attempts instead to acknowledge and embrace these generalization and integration issues from the start.

Let us finally return to the controversy at the opening of this article. We raised the question of what is to blame for the current concerns and challenges facing psychology? Would it be the methods and common approaches or the variable and complex nature of psychological phenomena? We believe that the best solution for these issues will come from approaches or combinations thereof (whatever they are) that lead us to a better understanding of psychological phenomena. This would require an open attitude in the awareness of different possible options. In this manuscript we have formulated and argued for one option, a perspective

on constructs that has not been sufficiently explicated and promoted in the literature to resolve the issues tied to heterogeneity in results and other dissatisfactions. It implies that many issues can be resolved by accepting that there is more variability and uncertainty in empirical findings than can be eliminated. In other words, variability is not merely due to sampling that can be reduced through increased sample size. Variability is inherent in the constructs and phenomena under study. The variability and uncertainty can be accommodated by constructs that are (rather fuzzy) areas and by using (and developing) methods that better assess uncertainty and generalization.

Although our proposal might intuitively seem to contradict the necessary condition of rigor, it in fact does not. Our proposal rather avoids rigidity and seeks realism in applying rigor. Uncertainties and sources thereof should not be ignored but recognized and assessed instead. The way forward with constructs will in fact follow from further research and will be subject of further developments that hopefully will lead to better generalization and integration.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

ORCID

Paul De Boeck  <http://orcid.org/0000-0002-0884-2582>

References

- AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.
- 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. *European Journal of Personality*, 27(2), 108–119. <https://doi.org/10.1002/per.1919>
- Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(4), 495–508. <https://doi.org/10.1080/10705511.2014.919210>
- Bardini, G., Politano, G., Benso, A., & Di Carlo, S. (2017). Multi-level and hybrid modelling approaches for systems biology. *Computational and Structural Biotechnology Journal*, 15, 396–402. <https://doi.org/10.1016/j.csbj.2017.07.005>
- Baribault, B., Donkin, C., Little, D. R., Trueblood, J. S., Oravecz, Z., van Ravenzwaaij, D., White, C. N., De Boeck, P., & Vandekerckhove, J. (2018). Meta-studies for robust tests of theory. *Proceedings of the National Academy of Sciences of the United States of America*, 115(11), 2607–2612. <https://doi.org/10.1073/pnas.1708285114>
- Barrett, L. F. (2017a). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, 12(1), 1–23. <https://doi.org/10.1093/scan/nsw154>
- Barrett, L. F. (2017b). *How emotions are made: The secret life the brain*. Houghton-Mifflin-Harcourt.
- Barry, A. E., Chaney, B., Piazza-Gardner, A. K., & Chavarria, E. A. (2014). Validity and reliability reporting practices in the field of health education and behavior: A review of seven journals. *Health Education & Behavior*, 41(1), 12–18. <https://doi.org/10.1177/1090198113483139>
- Beauchaine, T. P., Webster-Stratton, C., & Reid, M. J. (2005). Mediators, moderators, and predictors of one-year outcomes among children treated for early-onset conduct problems: A latent growth curve analysis. *Journal of Consulting and Clinical Psychology*, 73(3), 371–388. <https://doi.org/10.1037/0022-006X.73.3.371>
- Beauchaine, T. P., Zisner, A., & Sauder, C. L. (2017). Trait impulsivity and the externalizing spectrum. *Annual Review of Clinical Psychology*, 13(1), 343–368. <https://doi.org/10.1146/annurev-clinpsy-021815-093253>
- Blalock, H. M. (Ed.). (1971). *Causal models in the social sciences*. Aldine.
- Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley.
- Borsboom, D. (2005). *Measuring the mind. Conceptual issues in contemporary psychometrics*. Cambridge University Press.
- Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, 16(1), 5–13. <https://doi.org/10.1002/wps.20375>
- Borsboom, D., Deserno, M. K., Rhemtulla, M., Epskamp, S., Fried, E. I., McNally, R. J., Robinaugh, D. J., Perugini, M., Dalege, J., Constantini, G., Isvoranu, A.-M., Wysocki, A. C., van Borkulo, C. D., van Bork, R., & Waldorp, L. (2021). Network analysis of multivariate data in psychological science. *Nature Review Methods Primers*, 1, 58. <https://doi.org/10.1038/s43586-021-00055-w>
- Boschloo, L., van Borkulo, C. D., Borsboom, D., & Schoevers, R. A. (2016). A prospective study on how symptoms in a network predict the onset of depression. *Psychotherapy and Psychosomatics*, 85(3), 183–184. <https://doi.org/10.1159/000442001>
- Bridgman, P. W. (1927). *The logic of modern physics*. Macmillan.
- Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, 57(3), 153–178. <https://doi.org/10.1016/j.cogpsych.2007.12.002>
- Bukumiric, Z., Starcevic, V., Stanisavljevic, D., Marinkovic, J., Milic, N., Djukic-Dejanovic, S., Janjic, V., Corac, A., Ilic, A., Kostic, M., Nikolic, I., & Trajkovic, G. (2016). Meta-analysis of the changes in correlations between depression instruments used in longitudinal studies. *Journal of Affective Disorders*, 190, 733–743. <https://doi.org/10.1016/j.jad.2015.10.054>
- Burns, L., Preszler, J., Ahnach, A., Servera, M., & Becker, S. (2022). Multisource network and latent variable models of sluggish cognitive tempo, ADHD-inattentive, and depressive symptoms with Spanish children: Equivalent findings and recommendations. *Research on Child and Adolescent Psychopathology*, 50(7), 881–894. <https://doi.org/10.1007/s10802-021-00890-1>
- Bussemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, 100(3), 432–459. <https://doi.org/10.1037/0033-295X.100.3.432>
- Buss, D. M., & Craik, K. H. (1983). The act frequency approach to personality. *Psychological Review*, 90(2), 105–126. <https://doi.org/10.1037/0033-295X.90.2.105>
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.
- Cartwright, N., & Bradburn, N. (2011). A theory of measurement. In R. M. Li (Ed.), *The importance of common metrics for advancing social science theory and research: A workshop summary* (pp. 53–56). National Academies. <https://action.www.dur.ac.uk/chess/chessworkingpapers/>
- Cartwright, N., & Runhardt, R. (2014). Measurement. In N. Cartwright & E. Montuschi (Eds.), *Philosophy of social science: A new introduction* (pp. 265–287). Oxford University Press.
- Chambon, M., Dalege, J., Elberse, J. E., & van Harreveld, F. (2022). A psychological network approach to attitudes and preventive behaviors during pandemics: A COVID-19 study in the United Kingdom and the Netherlands. *Social Psychological and Personality Science*, 13(1), 233–245. <https://doi.org/10.1177/19485506211002420>
- Chang, H. (2021). Operationalism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Stanford University. <https://plato.stanford.edu/archives/fall2021/entries/operationalism>

- Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment*, 31(12), 1412–1427. <https://doi.org/10.1037/pas0000626>
- Clarke, D. D. (1987). Fundamental problems with fundamental research: A meta-theory for social psychology. *Philosophica*, 40(0), 23–61. <https://doi.org/10.21825/philosophica.82482>
- Crandall, C. S., & Sherman, J. W. (2016). On the scientific superiority of conceptual replications for scientific progress. *Journal of Experimental Social Psychology*, 66, 93–99. <https://doi.org/10.1016/j.jesp.2015.10.002>
- Crites, S., Fabrigar, L., & Petty, R. E. (1994). Measuring the affective and cognitive properties of attitudes: Conceptual and methodological issues. *Personality and Social Psychology Bulletin*, 20(6), 619–634. <https://doi.org/10.1177/0146167294206001>
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Harcourt Brace Javanovich.
- Cronbach, L. J. (1982). In praise of uncertainty. *New Directions for Program Evaluation*, 1982(15), 49–58. <https://doi.org/10.1002/ev.1310>
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302. <https://doi.org/10.1037/h0040957>
- Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. (1963). Theory of generalizability: A liberalization of reliability theory. *British Journal of Statistical Psychology*, 16(2), 137–163. <https://doi.org/10.1111/j.2044-8317>
- Cummins, R. (2010). *The world in the head*. Oxford Scholarship Online. <https://doi.org/10.1093/acprof:osobl/9780199548033.001.0001>
- Cureton, E. E. (1951). Validity. In E. F. Lindquist (Ed.), *Educational measurement* (pp. 621–694). American Council on Education.
- Dalege, J., Borsboom, D., van Harreveld, F., van den Berg, H., Conner, M., & van der Maas, H. L. (2016). Toward a formalized account of attitudes: The casual attitude network (CAN) model. *Psychological Review*, 123(1), 2–22. <https://doi.org/10.1037/a0039802>
- De Boeck, P., & Jeon, M. (2018). Perceived crisis and reforms: Issues, explanations, and remedies. *Psychological Bulletin*, 144(7), 757–777. <https://doi.org/10.1037/bul0000154>
- DeKay, M., Rubinchik, N., Li, Z., & De Boeck, P. (2022). Accelerating psychological science with metastudies: A demonstration using the risky-choice framing effect. *Perspectives on Psychological Science*, 17(6), 1704–1736. <https://doi.org/10.1177/17456916221079611>
- Doob, L. W. (1947). The behavior of attitudes. *Psychological Review*, 54(3), 135–156. <https://doi.org/10.1037/h0058371>
- Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Harcourt Brace Jovanovich.
- Edwards, J. R. (2011). The fallacy of formative models. *Organizational Research Methods*, 14(2), 370–388. <https://doi.org/10.1177/1094428110378369>
- Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, 5(2), 155–174. <https://doi.org/10.1037/1082-989X.5.2.155>
- Epstein, S. (1983). Aggregation and beyond: Some basic issues on the prediction of behavior. *Journal of Personality*, 51(3), 360–392. <https://doi.org/10.1111/j.1467-6494>
- Evans, N. J., Holmes, W. S., & Trueblood, J. S. (2019). Response-time data provide critical constraints on dynamic models of multi-alternative, multi-attribute choice. *Psychonomic Bulletin & Review*, 26(3), 901–933. <https://doi.org/10.3758/s13423-018-1557-z>
- Fabrigar, L. R., & Wegener, D. T. (2016). Conceptualizing and evaluating the replication of research results. *Journal of Experimental Social Psychology*, 66, 68–80. <https://doi.org/10.1016/j.jesp.2015.07.009>
- Fabrigar, L. R., Wegener, D. T., & Petty, R. E. (2020). A validity-based framework for understanding replication in psychology. *Personality and Social Psychology Review*, 24(4), 316–344. <https://doi.org/10.1177/1088868320931366>
- Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, 69(6), 1013–1027. <https://doi.org/10.1037/0022-3514.69.6.1013>
- Feest, U. (2005). Operationism in psychology: What the debate is about, what the debate should be about. *Journal of the History of the Behavioral Sciences*, 41(2), 131–149. <https://doi.org/10.1002/jhbs.20079>
- Feigl, H. (1950). Existential hypotheses. Realistic versus phenomenalist interpretations. *Philosophy of Science*, 17(1), 35–62. <https://doi.org/10.1086/287065>
- Flake, J. (2021). Strengthening the foundation of educational psychology by integrating construct validation into open science reform. *Educational Psychologist*, 56(2), 132–141. <https://doi.org/10.1080/00461520.2021.1898962>
- Flake, J., & 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. <https://doi.org/10.1177/2515245920952393>
- Flake, J. K., Pek, J., & Hehman, A. (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>
- Fried, E. I. (2017a). The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *Journal of Affective Disorders*, 208, 191–197. <https://doi.org/10.1016/j.jad.2016.10.019>
- Fried, E. I. (2017b). What are psychological constructs? On the nature and statistical modelling of emotions, intelligence, personality traits and mental disorders. *Health Psychology Review*, 11(2), 130–134. <https://doi.org/10.1080/17437199.2017.1306718>
- Fried, E. I. (2020). Theories and models: What they are, what they are for, and what they are about. *Psychological Inquiry*, 31(4), 336–344. <https://doi.org/10.1080/1047840X.2020.1854011>
- Fried, E. I., & Flake, J. K. (2018). Measurement matters. *APS Observer*, 31(3), 29–31.
- Fried, E. I., van Borkulo, C. D., Epskamp, S., Schoevers, R. A., Tuerlinckx, F., & Borsboom, D. (2016). Measuring depression over time ... Or not? Lack of unidimensionality and longitudinal measurement invariance in four common rating scales of depression. *Psychological Assessment*, 28(11), 1354–1367. <https://doi.org/10.1037/pas0000275>
- Funder, D. C., Levine, J. M., Mackie, D. M., Morf, C. C., Sansone, C., Vazire, S., & West, S. G. (2014). Improving the dependability of research in personality and social psychology: Recommendations for research and educational practice. *Personality and Social Psychology Review*, 18(1), 3–12. <https://doi.org/10.1177/1088868313507536>
- Greenwald, A., & Lai, C. (2020). Implicit social cognition. *Annual Review of Psychology*, 71(1), 419–445. <https://doi.org/10.1146/annurev-psych-010419-050837>
- 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. <https://doi.org/10.1037/0022-3514.74.6.1464>
- Gross, C. T., & Canteras, N. S. (2012). The many paths to fear. *Nature Reviews. Neuroscience*, 13(9), 651–658. <https://doi.org/10.1038/nrn3301>
- Hofstee, W. K., de Raad, B., & Goldberg, L. R. (1992). Integration of the Big Five and circumplex approaches to trait structure. *Journal of Personality and Social Psychology*, 63(1), 146–163. <https://doi.org/10.1037/0022-3514.63.1.146>
- Hull, C. L. (1943). *Principles of behavior: An introduction to behavior theory*. Appleton-Century-Crofts.
- Huys, Q. J. M., Maia, T. V., & Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature Neuroscience*, 19(3), 404–413. <https://doi.org/10.1038/nn.4238>
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemaľcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., Hasselman, F., Hicks, J. A., ... Nosek, B. A. (2014). Investigating variation in replicability. A “Many Labs” replication project. *Social Psychology*, 45(3), 142–152. <https://doi.org/10.1027/1864-9335/a000178>

- Koch, S. (1992). Psychology's Bridgman and Bridgman's Bridgman. An essay in reconstruction. *Theory & Psychology*, 2(3), 261–290. <https://doi.org/10.1177/0959354392023002>
- Krajich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based choice. *Proceedings of the National Academy of Sciences*, 108(33), 13852–13857. <https://doi.org/10.1073/pnas.1101328108>
- Kretzschmar, A., & Nebe, S. (2021). Working memory, fluid reasoning, and complex problem solving: Different results explained by the Brunswick symmetry. *Journal of Intelligence*, 9(1), 5. <https://doi.org/10.3390/jintelligence9010005>
- Land, K. C. (1970). On the estimation of path coefficients for unmeasured variables from correlations among observed variables. *Social Forces*, 48(4), 506–511. <https://doi.org/10.2307/2575575>
- Landry, O., & Chouinard, P. A. (2016). Why we should study the broader autism phenotype in typically developing populations. *Journal of Cognition and Development*, 17(4), 584–595. <https://doi.org/10.1080/15248372.2016.1200046>
- Lee, M. D., Criss, A. H., Devezer, B., Donkin, C., Etz, A., Leite, F. P., Matzke, D., Rouder, J. N., Trueblood, J. S., White, C. N., & Vandekerckhove, J. (2019). Robust modeling in cognitive science. *Computational Brain & Behavior*, 2(3–4), 141–153. <https://doi.org/10.1007/s42113-019-00029-y>
- Leitgöb, H., Seddig, D., Asparouhov, T., Behr, D., Davidov, E., De Roover, K., Jak, S., Meitinger, K., Menold, N., Muthén, B., Rudnev, M., Schmidt, P., & van de Schoot, R. (2023). Measurement invariance in the social sciences: Historical development, methodological challenges, state of the art, and future perspectives. *Social Science Research*, 110, 102805. <https://doi.org/10.1016/j.ssresearch.2022.102805>
- Lerche, V., & Voss, A. (2017). Retest reliability of the parameters of the Ratcliff diffusion model. *Psychological Research*, 81(3), 629–652. <https://doi.org/10.1007/s00426-016-0770-5>
- Lerche, V., von Krause, M., Voss, A., Frischkorn, G. T., Schubert, A.-L., & Hagemann, D. (2020). Diffusion modeling and intelligence: Drift rates show both domain-general and domain-specific relations with intelligence. *Journal of Experimental Psychology: General*, 149(12), 2207–2249. <https://doi.org/10.1037/xge0000774>
- Lewin, K. (1942). Field theory and learning. In N. B. Henry (Ed.), *The forty-first yearbook of the National Society for the Study of Education: Part 2, The psychology of learning* (pp. 215–242). University of Chicago Press. <https://doi.org/10.1037/11335-006>
- Lewin, K. (1943). Defining the 'field at a given time'. *Psychological Review*, 50(3), 292–310. <https://doi.org/10.1037/h0062738>
- 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>
- Lissitz, R. W., & Samuelsen, K. (2007). A suggested change in terminology and emphasis regarding validity and education. *Educational Researcher*, 36(8), 437–448. <https://doi.org/10.3102/0013189X07311286>
- Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, 3(7), 635–694. <https://doi.org/10.2466/PRO.3.7.635-694>
- Lord, C., Charman, T., Havdahl, A., Carbone, P., Anagnostou, E., Boyd, B., Carr, T., de Vries, P. J., Dissanayake, C., Divan, G., Freitag, C. M., Gotelli, M. M., Kasari, C., Knapp, M., Mundy, P., Plank, A., Scahill, L., Servili, C., Shattuck, P., ... McCauley, J. B. (2022). The Lancet Commission on the future of care and clinical research in autism. *Lancet*, 399(10321), 271–334. [https://doi.org/10.1016/S0140-6736\(21\)01541-5](https://doi.org/10.1016/S0140-6736(21)01541-5)
- MacCorquodale, K., & Meehl, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review*, 55(2), 95–107. <https://doi.org/10.1037/h0056029>
- Machery, E. (2020). What is a replication? *Philosophy of Science*, 87(4), 545–567. <https://doi.org/10.1086/709701>
- Machery, E. (2021). A mistaken confidence in data. *European Journal for Philosophy of Science*, 11(2), 34. <https://doi.org/10.1007/s13194-021-00354-9>
- MacKenzie, S. B. (2003). The dangers of poor construct conceptualization. *Journal of the Academy of Marketing Science*, 31(3), 323–326. <https://doi.org/10.1177/0092070303031003011>
- Mandelbrot, B. (1983). *The fractal geometry of nature*. Freeman. <https://doi.org/10.2307/2323761>
- Mari, L., Wilson, M., & Maul, A. (2019). *Measurement across the sciences*. Springer. <https://doi.org/10.1007/978-3-030-65558-7>
- McPhetres, J., Albayrak-Aydemir, N., Barbosa Mendes, A., Chow, E. C., Gonzalez-Marquez, P., Loukras, E., Maus, A., O'Mahony, A., Pomareda, C., Primbs, M. A., Sackman, S. L., Smithson, C. J. R., & Volodko, K. (2021). A decade of theory as reflected in Psychological Science (2009–2019). *PLoS One*, 16(3), e0247986. <https://doi.org/10.1371/journal.pone.0247986>
- McShane, B., Tackett, J., Böckenholt, U., & Gelman, A. (2019). Large-scale replication projects in contemporary psychological research. *The American Statistician*, 73(sup1), 99–105. <https://doi.org/10.1080/00031305.2018.1505655>
- Merkle, E. C., & Van Zandt, T. (2006). An application of the Poisson race model to confidence calibration. *Journal of Experimental Psychology: General*, 135(3), 391–408. <https://doi.org/10.1037/0096-3445.135.3.391>
- Milkman, K. L., Patel, M. S., Gandhi, L., Graci, H. N., Gromet, D. M., Ho, H., Kay, J. S., Lee, T. W., Akinola, M., Beshears, J., Bogard, J. E., Buttenheim, A., Chabris, C. F., Chapman, G. B., Choi, J. J., Dai, H., Fox, C. R., Goren, A., Hilchey, M. D., ... Duckworth, A. L. (2021). A megastudy of text-based nudges encouraging patients to get vaccinated at an upcoming doctor's appointment. *Proceedings of the National Academy of Sciences*, 118(20), e2101165118. <https://doi.org/10.1073/pnas.2101165118>
- Miller, A. H., & Raison, C. (2016). The role of inflammation in depression: From evolutionary imperative to modern treatment target. *Nature Reviews. Immunology*, 16(1), 22–34. <https://doi.org/10.1038/nri.2015.5>
- Mischel, W. (1968). *Personality and assessment*. Wiley.
- Möttus, R., Kandler, C., Bleidorn, W., Riemann, R., & McCrae, R. R. (2017). Personality traits below facets: The consensual validity, longitudinal stability, heritability, and utility of personality nuances. *Journal of Personality and Social Psychology*, 112(3), 474–490. <https://doi.org/10.1037/pspp0000100>
- Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, 17(3), 313–335. <https://doi.org/10.1037/a0026802>
- Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, 104(2), 266–300. <https://doi.org/10.1037/0033-295X.104.2.266>
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- O'Connor, T. G., Deater-Deckard, K., Fulker, D., Rutter, M., & Plomin, R. (1998). Genotype-environment correlations in late childhood and early adolescence: Antisocial behavioral problems and coercive parenting. *Developmental Psychology*, 34(5), 970–981. <https://doi.org/10.1037/0012-1649.34.5.970>
- O'Hagan, T. (2004). Dicing with the unknown. *Significance*, 1(3), 132–133. <https://doi.org/10.1111/j.1740-9713.2004.00050.x>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- Osborne, J. W., & Fitzpatrick, D. C. (2012). Replication analysis in exploratory factor analysis: What it is and why it makes your analysis better. *Practical Assessment, Research and Evaluation*, 17, 15. <https://doi.org/10.7275/h0bd-4d11>
- Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, 7(6), 531–536. <https://doi.org/10.1177/1745691612463401>
- Patterson, G. R. (1986). Performance models for antisocial boys. *The American Psychologist*, 41(4), 432–444. <https://doi.org/10.1037/0003-066X.41.4.432>
- Paunonen, S. V., & Ashton, M. C. (2001). Big Five factors and facets and the prediction of behavior. *Journal of Personality and Social*

- Psychology*, 81(3), 524–539. <https://doi.org/10.1037/0022-3514.81.3.524>
- Pek, J., & Park, J. (2019). Complexities in power analysis: Quantifying uncertainties with a Bayesian-classical hybrid approach. *Psychological Methods*, 24(5), 590–605. <https://doi.org/10.1037/met0000208>
- Pek, J., Pitt, M., & Wegener, D. (2022). Uncertainty limits the use of power analysis. *Journal of Experimental Psychology General*. <https://doi.org/10.1037/xge0001273>
- Petty, R. E. (2018). The importance of exact conceptual replications. *The Behavioral and Brain Sciences*, 41, e146. <https://doi.org/10.1017/S0140525X18000821>
- Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, 19, 123–205. [https://doi.org/10.1016/S0065-2601\(08\)60214-2](https://doi.org/10.1016/S0065-2601(08)60214-2)
- Petty, R. E., & Cacioppo, J. T. (1981). *Attitudes and persuasion: Classic and contemporary approaches*. Wm. C. Brown.
- Petty, R. E., & Wegener, D. T. (1998). Attitude change: Multiple roles for persuasion variables. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., pp. 323–390). McGraw-Hill.
- Proshansky, J. A. (1943). A projective method for the study of attitudes. *The Journal of Abnormal and Social Psychology*, 38(3), 393–395. <https://doi.org/10.1037/h0056930>
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2), 59–108. <https://doi.org/10.1037/0033-295X.85.2.59>
- Rips, L. J., & Conrad, F. G. (1989). Folk psychology of mental activities. *Psychological Review*, 96(2), 187–207. <https://doi.org/10.1037/0033-295X.96.2.187>
- Robitzsch, A., & Lüdtke, O. (2022). *Why measurement invariance is not necessary for valid group comparisons*. PsyArXiv, <https://doi.org/10.31234/osf.io/cjyqp>
- Schmittmann, V. D., Cramer, A. O. J., Waldorp, L. J., Epskamp, S., Kievit, R. A., & Borsboom, D. (2013). Deconstructing the construct: A network perspective on psychological phenomena. *New Ideas in Psychology*, 31(1), 43–53. <https://doi.org/10.1016/j.newideapsych.2011.02.007>
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton-Mifflin.
- Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, 9(1), 76–80. <https://doi.org/10.1177/1745691613514755>
- 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. <https://doi.org/10.1177/1745691617708630>
- Singer, A., Lutz, A., Escher, J., & Halladay, A. (2023). A full semantic toolbox is essential for autism research and practice to thrive. *Autism Research: official Journal of the International Society for Autism Research*, 16(3), 497–501. <https://doi.org/10.1002/aur.2876>
- Skinner, B. F. (1963). Operant behavior. *American Psychologist*, 18(8), 503–515. <https://doi.org/10.1037/h0045185>
- Slaney, K. L. (2017). *Validating psychological constructs: Historical, philosophical, and practical dimensions*. Palgrave Macmillan.
- Slaney, K. L., & Garcia, D. A. (2015). Constructing psychological objects: The rhetoric of constructs. *Journal of Theoretical and Philosophical Psychology*, 35(4), 244–259. <https://doi.org/10.1037/teo0000025>
- Smith, G. T. (2005). On construct validity: Issues of method and measurement. *Psychological Assessment*, 17(4), 396–408. <https://doi.org/10.1037/1040-3590.17.4.396>
- Stanley, T. D., Carter, E. C., & Doucouliagos, H. (2018). What meta-analyses reveal about the replicability of psychological research. *Psychological Bulletin*, 144(12), 1325–1346. <https://doi.org/10.1037/bul0000169>
- Sternberg, D., & Detterman, R. (Eds.), (1986). *What is intelligence? Contemporary viewpoints on its nature and definition*. Ablex.
- Stevens, S. S. (1935). Kret operational definition of psychological concepts. *Psychological Review*, 42(6), 517–527. <https://doi.org/10.1037/h0056973>
- Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science*, 9(1), 59–71. <https://doi.org/10.1177/1745691613514450>
- Sucksmith, E., Roth, I., & Hoekstra, R. (2011). Autistic traits below the clinical threshold: Re-examining the broader autism phenotype in the 21st century. *Neuropsychology Review*, 21(4), 360–389. <https://doi.org/10.1007/s11065-011-9183-9>
- Sudol, K., & Mann, J. J. (2017). Biomarkers of suicide attempt behavior: Towards a biological model of risk. *Current Psychiatry Reports*, 19(6), 31. <https://doi.org/10.1007/s11920-017-0781-y>
- Tal, E. (2020). Measurement in science. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Stanford University. <https://plato.stanford.edu/archives/fall2020/entries/measurement-science/>
- Tolman, E. C. (1932). *Purposive behavior in animals and men*. Appleton-Century.
- Tolman, E. C. (1936). An operational analysis of “demands”. *Erkenntnis*, 6, 383–392. <http://www.jstor.org/stable/20011843>
- Tolman, E. C. (1938). The determiners of behavior at a choice point. *Psychological Review*, 45(1), 1–41. <https://doi.org/10.1037/h0062733>
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, 108(3), 550–592. <https://doi.org/10.1037/0033-295X.108.3.550>
- VanderWeele, T. J. (2022). Constructed measures and causal inference: Towards a new model of measurement for psychosocial constructs. *Epidemiology*, 33(1), 141–151. <https://doi.org/10.1097/EDE.0000000000001434>
- Wallace, L. E., Patton, K. M., Luttrell, A., Sawicki, V., Fabrigar, L. R., Teeny, J., MacDonald, T. K., Petty, R. E., & Wegener, D. T. (2020). Perceived knowledge moderates the relation between subjective ambivalence and the “impact” of attitudes: An attitude strength perspective. *Personality & Social Psychology Bulletin*, 46(5), 709–722. <https://doi.org/10.1177/0146167219873492>
- Wegener, D. T., & Fabrigar, L. R. (2004). Constructing and evaluating quantitative measures for social psychological research. Conceptual challenges and methodological solutions. In C. Sansone, C.C. Morf, & A.T. Panter (Eds.), *The Sage handbook of methods in social psychology* (pp. 145–172). Sage.
- Wittmann, W. W. (1988). Multivariate reliability theory. Principles of symmetry and successful validation strategies. In J. R. Nesselrode & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 505–560). Plenum Press.
- Wu, H., & Browne, M. W. (2015). Quantifying adventitious error in a covariance structure as a random effect. *Psychometrika*, 80(3), 571–600. <https://doi.org/10.1007/s11336-015-9451-3>
- Zmigrod, L., Eisenberg, I. W., Bissett, P. G., Robbins, T. W., & Poldrack, R. A. (2021). The cognitive and perceptual correlates of ideological attitudes: A data-driven approach. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 376(1822), 20200424. Article 20200424. <https://doi.org/10.1098/rstb.2020.0424>
- Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, 41, e120. <https://doi.org/10.1017/S0140525X17001972>