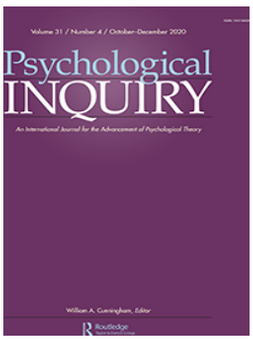

Routledge
Taylor & Francis Group

Psychological Inquiry

An International Journal for the Advancement of Psychological Theory

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hpli20

Theories and Models: What They Are, What They Are for, and What They Are About

Eiko I. Fried

To cite this article: Eiko I. Fried (2020) Theories and Models: What They Are, What They Are for, and What They Are About, Psychological Inquiry, 31:4, 336-344, DOI: 10.1080/1047840X.2020.1854011

To link to this article: https://doi.org/10.1080/1047840X.2020.1854011

PSYCHOLOGICAL INQUIRY
2020, VOL. 31, NO. 4, 336–344
https://doi.org/10.1080/1047840X.2020.1854011

Routledge
Taylor & Francis Group

REPLY

Theories and Models: What They Are, What They Are for, and What They Are About

Eiko I. Fried

Department of Clinical Psychology, Leiden University, Leiden, The Netherlands

In Fried (this issue), I argue that a lot of work in the social sciences generally—and in psychology specifically—reads like an exercise in statistical model fitting, and falls short of theory building and testing in three ways. First, theories are absent, which fosters conflating statistical models with theoretical models. Second, theories are latent, that is, implied but not explicated. Third, theories are weak, that is, ambiguous and impossible to test or reject because they fit any data. I focus on psychometric factor and network models and their applications to cognitive, personality, and clinical psychology, showing that selecting statistical models that impose assumptions consistent with the theories they are supposed to corroborate is necessary for bringing data to bear on these theories.

Seven commentaries agree with some of the core challenges the field faces. They raise some important criticisms of the target article and provide extensions by identifying further problems and potential solutions. Here, I aim to integrate some of the core points and criticism raised, and provide a brief primer on theory formation, structured into three sections: (1) what are theories; (2) what are theories for; (3) and what are theories about. This is followed by a section dedicated to the question (4) how to develop theories. I conclude with (5) specific obstacles to theory formation psychological scientists face, and how they can be overcome.

What Are Theories

The terms *theory* and *model* have been defined in numerous ways, and there are at least as many ideas on how theories and models relate to each other (Bailer-Jones, 2009). I understand *theories* as bodies of knowledge that are broad in scope and aim to explain robust phenomena. *Models*, on the other hand, are instantiations of theories, narrower in scope and often more concrete, commonly applied to a particular aspect of a given theory, providing a more local description or understanding of a phenomenon. Evolution and gravitation are theories, but their application to mate selection and the motion of the planets is best done via models. From this perspective, models serve as intermediaries between theories and the real world.

I agree with DeYoung & Krueger (this issue) who describe good theories as coherent and useful conceptual frameworks into which existing knowledge can be integrated. These bodies of knowledge clarify not only the targets of research but also help distill the right kinds of theoretical and empirical questions. Good theories should provide better *abductive explanations* about the phenomena under investigation than competing theories. Abductive inference, or inference to the best explanation, can be understood as a process that aims to provide good (i.e., simple and plausible) explanations for phenomena, in the sense that if a theory were true, the phenomenon would look the way it looks in the world (Peirce, 1931).

In contrast to theories, models—as highlighted by Smaldino (this issue)—illustrate with precision the mechanisms that might govern the processes that lead to a phenomenon by decomposing processes into relevant parts, properties of these parts, relations between the parts, and temporal dynamics of their change (cf. Smaldino, 2020). Smaldino describes models using the metaphor of geographical maps that are useful *because* they are simplified: models ignore much of reality to be useful, in the same way, a map of Rome ignores much of reality to help us navigate Rome, a point we will return to in detail later. The tricky part is to leave out the right kinds of things to enable models to help with explanation or prediction. This is challenging because it amounts to "a reckoning with one's conceptualization of reality" rather than a simple algorithm (Smaldino, this issue).

In the target article, I separate theoretical models (i.e., theories) from statistical models (e.g., structural equation and network models), and discuss their relations. But the broader class of models can take many forms, such as verbal or mathematical models (e.g., different descriptions of how neurons interact in the brain), diagrams (e.g., the hydrological cycle that moves water on Earth), or mechanical models (e.g., a physical model of the DNA double helix) (Bailer-Jones, 2009). In psychology, such models are *phenomena models* because they usually aim to represent robust phenomena in the world, such as the phenomenon that mental disorders are often comorbid. Statistical models, on the other hand, are *data models* because they provide representations of data (Haslbeck, Ryan, Robinaugh, & Waldorp, 2019). This provides a glimpse of the problem psychology is facing: common data models are typically far removed from concrete instantiations of theories that provide a bridge to

CONTACT Eiko I. Fried www.eiko-fried.com; eikofried@gmail.com Department of Clinical Psychology, Leiden University, Leiden, The Netherlands.

the real world, and data models are regularly interpreted as theoretical models. This inference gap makes it challenging to bring data to bear on theories.

DeYoung & Krueger (this issue) summarize this inference gap: hoping to "determine causal relations merely by fitting a statistical model […] in typical cross-sectional data is wishing for the impossible," (p. 291) and it is "imperative for researchers to remember that paths in the structural equation or network models are not inherently causal" (p. 291). This is true, but these statements are in contrast to Krueger's foundational work for the literature critically discussed in the target article. In an analysis of cross-sectional data, he defines structural equation models as "a means of *explaining* the correlations […] among variables […] by postulating that these correlations arise because of the *influence* of a smaller number of underlying, latent dimensions," and concludes in the abstract that "comorbidity *results* from common, underlying core psychopathological processes" (my highlights) (Krueger, 1999). Granted, this was over 2 decades ago, but it is to show that DeYoung & Krueger (this issue) provide a welcome and critical reminder all psychologists should heed.

What Are Theories for?

The target article conceptualizes psychological theories on a dimension between weak and strong theories and characterizes stronger theories as non-ambiguous and clearly explicated theories that make testable (and in the best case precise) predictions. Much in the same way, Conway et al. (this issue) argue that good theories generate testable predictions, and Watts et al. (this issue) posit that theories in psychology such as explanations of the *p* factor of psychopathology ought to be put to more rigorous tests. This focus on testing precise predictions generated by theories brings into focus the question of what theories are *for* and contrasts with other views expressed that highlight the value of theories for understanding (Van Rooij & Baggio, this issue) or utility (Yarkoni, this issue). This diversity of perspectives is not unique to psychology, and I recommend the interview with scientists from various fields that Bailer-Jones conducted on this topic (Bailer-Jones, 2009).

Testing, Testing, Testing

Psychology has focused strongly on theories to make *testable predictions*. This, I believe, is born in part out of frustration that theories have often been inherently untestable. My first scientific project was concerned with translating common evolutionary theories of depression into agent-based models to determine under which conditions potential adaptive benefits of depression could arise (Nettle, 2004). But leading theories were weak theories: narrative and ambiguous. Most theories, in fact, did not even specify what they meant by the phenomenon of depression that they aimed to explain (increased levels of sad mood on the one hand, or recurrent, severe, clinical depression on the other). But how to translate theories into a model when central parts are not sufficiently explicated? Whether depression came out as adaptive or not in my agent-based models was dependent entirely on my speculation about what the theorists may have meant, and, as a result, had little bearing on said theories.

Watts et al. (this issue) and Conway et al. (this issue) demonstrate similar examples from other disciplines, where theories are either weak or conflated with statistical models. Note that I am not arguing that writing up initial ideas verbally is unimportant for theory formation; problems arise when we defend and reify such ideas for decades without exposing them to any serious risk of refutation. One way of putting theories at risk is testing, and one of the most visible proponents of risky tests in the psychological literature is Meehl (1978, 1990a). This focus on testing specific hypotheses is broadly consistent with the hypothetico-deductive framework under which most of the psychology operates: we derive hypotheses from theories, deduce observational consequences from these hypotheses, and test these consequences to verify the hypotheses (and thereby the theory) using the logic of confirmation (Borsboom, van der Maas, Dalege, Kievit, & Haig, 2020; Robinaugh, Haslbeck, Ryan, Fried, & Waldorp, 2020). Curiously, the hypothetico-deductive framework and its focus on testing have in part reinforced the lack of theory formation in psychology, because, like the work of Meehl himself, it provides little in the way of helping psychologists to construct and update theories after they are exposed to empirical data (Robinaugh et al., 2020). As Borsboom et al. (2020) put it: the framework "discourages the use of systematic methods in order to generate, develop, and fully appraise theories." The focus has led to the very consequence that Meehl bemoaned: that psychological theories "tend neither to be refuted nor corroborated, but instead merely fade away as people lose interest" (Meehl, 1978).

Theories Provide Explanations

Van Rooij & Baggio (this issue) remind us that theories are primarily *for explaining*: "testing is but one means, among others, of assessing, revising and refining theories, but this is a secondary research activity. One needs theory first to know what is worth testing" (p. 324). Like Cummins (2000), they stress that many theories in psychology are explananda, not explanantia, and themselves require explaining. Theories should therefore have *a priori* verisimilitude—they should, among others, be theoretically viable, sufficient to produce explananda, and internally consistent before they are submitted to testing. This focus on theory formation would have made my task to translate evolutionary theories of depression into formal models a lot easier.

The call by van Rooij & Baggio (this issue) aligns with recent work arguing that hypothesis testers should spend less time testing hypotheses—because coming up with informative hypotheses *worth testing* requires theoretical work first (Scheel et al., 2020). Such theoretical work, the authors argue, does often not take place, which can be seen in the facts that psychologists often struggle to precisely write down hypotheses or specify the smallest effect size of interest. If theories are for understanding, then failing to do

the theoretical legwork before testing will lead to arbitrary tests with arbitrary inferences.

Theories Should be Useful

Yarkoni (this issue) provides a third perspective of pragmatism, arguing that the distinction between weak and strong theories is largely in the eye of the beholder. A good theory is not necessarily one that is formalized or precise, but one that is more *useful for a certain purpose*, such as making accurate predictions and contributing to objective progress. He then argues that strong or formal theories, for example, in clinical psychology, do not have a strong track record of being useful (e.g., in informing public health policies or clinical practice), and that the things that work best in clinical psychology and psychiatry—pharmacological interventions—were developed with little regard to theory.

First, I am not aware of atheoretical work using artificial intelligence or machine learning that has led to dramatic policy shifts in clinical psychology or provided substantial improvements for broad areas of clinical care. Second, even if one were to agree, for the sake of the argument, that strong theories have a weak track record in psychology, they certainly do not have a weak track record in other sciences. It also seems premature to conclude that theorizing has failed us because the majority of psychology has not concerned itself with proper theory building in the first place (Borsboom, 2013). This is especially the case for clinical psychology: the target article does not provide evidence "that the use of formal theories in psychology has historically done much to advance clinical practice" (Yarkoni, this issue, p. 332) because there isn't exactly an abundance of formal theories on mental illness to begin with (Burger et al., 2020; Robinaugh et al., 2019). Finally, cognitive behavior therapy (CBT), embedded in a broad and powerful theoretical framework, has shown to be at least as effective as pharmacological interventions for a large range of mental disorders (cf. Barth et al., 2013; Cuijpers et al., 2020), and is the first-line treatment of choice for adolescents and adults in many countries (Hollon & Beck, 2013). I, therefore, disagree that "perhaps theory-building is actually not a very effective way for applied psychologists to make objective progress" (Yarkoni, this issue, p. 332).

Overall, Yarkoni's pragmatic appeal is a welcome reminder: we may also want to evaluate theories on whether they provide actionable outcomes (although I note that many of science's most powerful and successful theories did not initially provide such outcomes). This, one may argue, holds especially for disciplines where the core goal is to improve the lives of people, for example, with major depression. In this context, it may be most useful to determine what treatment works for whom. *Why* this is the case is not very relevant, a pragmatist might say, as long as we can assign people to the appropriate treatments reliably (cf. Yarkoni & Westfall, 2017). But there can be no doubt that proper *understanding* of a phenomenon is one of the best ways toward better prediction and control: "There is nothing more practical than a good theory" (Lewin, 1952), after all. Being able to predict with precision the regularity in disease progression of syphilis provided no levers for interventions—an understanding of the (in this case, biological) mechanisms that governed the disease system, however, ultimately led to developments of the first antibiotic in 1910. As I argue later, such insights into mechanisms and interventions also exist at the psychological level.

Vaidyanathan (this issue) highlights the pragmatic concern that different models may be more useful in certain contexts. I agree: whether a person requires treatment for depression may benefit from a different model than what particular treatment a person should receive, in much the same way that different types of maps for Rome maximize utility for different purposes. Such maps differ from each other in that they ignore different parts of Rome in order to be most useful (finding the closest café for a good espresso vs organizing a 2-hour community walk with the elderly that minimizes topological obstacles). Nonetheless, these different maps are all instantiations of the same underlying target system (Rome), maybe in the same way that different models of depression may be derived from a single overarching theory of depression.

Theories are for Description, Explanation, Prediction, and Control

In summary, I believe that theories are *for* helping us to understand the world, which facilitates description, prediction, and control. If we take the value of theories seriously, then we should develop better theoretical practices and put our theories at greater risk of refutation. This risk comes in many forms, not only in the form of testing. Risk comes from asking whether the theory is at all useful in helping us understand the world; from thinking clearly about what a theory is meant to explain; from having a clear theory in the first place; from thinking clearly about *a priori* verisimilitude, viability, coherence, breadth, and depth of a theory before it is exposed to data; from translating a theory into a formal model; from simulating data from a formal model and testing whether it matches the phenomena we observe in the world; and from deriving predictions from a theory and testing them in observational or experimental designs.

What Are Theories About?

After defining what theories *are* and what they are *for*, we will now turn to what psychological theories are *about*. In the target article, I posit that theories are *about phenomena*—robust and recurring features in the world. More specifically, theories seek to explain such phenomena (Apel, 2011; Woodward, 2011). In psychology, phenomena include that a considerable proportion of people will experience panic attacks in their lifetime; that subtests on intelligence scales are highly intercorrelated; and that most mental disorders do not seem to have simple, monocausal explanations. There are also other phenomena, for example, that there are many more p-values of around 4% in the published psychological literature than there are p-values of around 6%. I will not concern myself with such kinds of phenomena here but

focus on emergent psychological constructs such as traits, emotions, attributes, or mental disorders. The five points this section addresses are whether theories are about statistical effects or robust phenomena in the world; how theories represent target systems; at what levels such systems should be represented; challenges of questionable measurement practices; and the "realness" of phenomena that our theories seek to explain.

Phenomena vs Effects

Muthukrishna and Henrich (2019) whom I quote in the target article argue that robust and replicable effects can be understood as solid stones—but those sound theories are required to build the house. This invites the interpretation that theories are *about* effects. Van Rooij & Baggio (this issue) remind us that we "first need good candidate theories to guide us and to determine which statistical effects would be relevant at all for assessing, updating, revising and refining theory," (p. 323) and that "the idea that we can build a theory on theory-neutral 'facts' has been philosophically discredited long ago" (p. 323). I agree, and it may be useful to distinguish more clearly between *robust statistical effects* (in data) and *robust phenomena* (in the world). Both require explaining, but the latter is the solid stones that facilitate building proper theory houses.

From this perspective, psychology's response to the replicability crisis can be understood as an effort to distinguish robust effects from noise caused by sloppy work, questionable research practices, publication bias, and so on (cf. Robinaugh et al., 2020). But robust statistical effects do not by themselves establish robust phenomena (Apel, 2011; Woodward, 2011). For example, some mental health datasets feature extremely strong correlations among depression symptoms. Interestingly, the measures by which such data were acquired contain a skip structure: if participants do not endorse questions 1 or 2, questions 3 through 9 are missing by design. Researchers have often imputed these missing data with zeros, inducing strong spurious relations among items (Borsboom et al., 2017). The robust statistical effect of very large correlations is therefore explained by a common, questionable imputation strategy for missing data, and not a robust phenomenon that requires theory formation. I conclude that robust statistical effects are therefore neither sufficient (cf. zero imputation) nor necessary (cf. capacity for language) to establish robust phenomena, but they will more often than not help to discover and define them. Many robust phenomena are based on robust effects.

Theories Represent Target Systems

When it comes to the question of what theories are *about*, I share the view summarized in Haslbeck et al. (2019) that psychological constructs can be thought of as *target systems*, which in turn consist of components and relations among these components. Such target systems are *represented* via a theory's structure, which, like the target system, features components and relations among them (Haslbeck et al., 2019; Robinaugh et al., 2020). If the target system is well represented by the theory's structure, we can engage in *surrogative reasoning*, that is, understand the real-world target system from our theory, and predict the future behavior of the target system via our theory. Maps are most useful when their structure represents the target system (e.g., Rome) well. This view of a target *system* is important because psychological constructs such as mental disorders are results of complex biopsychosocial interactions (van Rooij & Baggio this issue, Smaldino this issue), and progress in understanding (and, consequently, predicting and treating) them will require studying the systems from which these constructs arise. Importantly, such systems interact with each other. For example, fully understanding a person's depression likely requires not only taking into account a person's physical health system and mental health system, but also their social system and resilience system, as well as the healthcare system, education system, and political system of the country they live in. As we argue elsewhere: it's systems all the way down (Fried & Robinaugh, 2020).

Is theory formation doomed to fail in psychology because theories are about such highly complex target systems? Yarkoni (this issue) states that theory-driven approaches will often be unable to approximate the complex data-generating mechanisms, greatly limiting their utility in faithfully describing systems and hence predicting important outcomes. Vaidyanathan (this issue) highlights that clinical and applied contexts do not have the luxury to wait until perfect theories emerge. And DeYoung & Krueger (this issue) argue that my target article asks psychology for a "focus on a kind of theory that it cannot generally be expected to produce at this time" (p. 289). I believe that we can learn most about our target systems when our theories and models *represent* them well, that is, with some level of abstraction. When we understand models as intermediaries between theories and the real world, the idea that all models are wrong because they are incomplete (Meehl, 1990a) is a feature, not a bug; a model's inability to perfectly reflect reality does not stand in the way of providing actionable insights, as shown by the Rutherford model or Newton's theory of gravitation.

A Problem of Levels

The framework in which a target system is represented by a theory's structure provides little general guidance on the right level at which components should be studied (e.g., specific symptoms vs higher-order traits such as the *p* factor), nor what the right level of explanation is (e.g., biological vs psychological vs social mechanisms). No matter what one proposes as an explanation for a phenomenon, others could always argue that a true explanation requires grappling with a lower level in the hierarchy: it's turtles all the way down. This, Yarkoni (this issue) argues, muddles the boundaries between explanation and description, because one researcher's explanation (e.g., at the psychological level) may be another researcher's description (some argue that true explanation takes place at the biological or physical level).

This challenge makes the distinction between explanans and explanandum arbitrary.

The argument ties into the importance of *mechanistic explanations* raised by van Rooij & Baggio (this issue). Suppose the goal of our theory is to explain the capacity of a modern coffee machine for producing coffee. This requires, first, to separate the coffee machine into its components and relations among these, and, second, to demonstrate how interactions among components give rise to the capacity (Van Rooij & Baggio, 2020). However, mechanistic explanations go one step further in that it is not sufficient that components and relations are described in a useful sketch (a Smaldinoesque map). It is required, thirdly, that they faithfully describe *physical realizations* of these components and relations, along with their properties. This is possible (and undoubtedly useful) for the coffee machine, for which we know the nature and properties of its parts (e.g., the alloy of particular components), as well as their relations. It offers levers for interventions, in case the machine breaks. But I wonder whether mechanistic explanations are feasible or necessary for psychological explananda.

These points raised by Yarkoni (this issue) and van Rooij & Baggio (this issue) can be framed as a problem of *levels*. Four arguments are worth noting. First, "no one would seek to understand the origin of hypertension at the level of quarks" (Kendler, 2005). That is, we can rule out some levels to provide informative explanations of the phenomena we are interested in. Second, the idea that lower (usually biological rather than psychological) levels in the hierarchy naturally provide better or more powerful explanations comes in many flavors and is known as biological or explanatory reductionism. This view often fails to provide sufficient explanations for emergent phenomena at higher levels. Eronen's primer on the levels problem (Eronen, 2019) lists many examples of such phenomena like the heartbeat or avian flight, concluding that "explaining or predicting the behavior of complex systems requires higher-level explanations," and that "in this light, it is perhaps not surprising that reductionistic research programs in psychiatry that target lower biological levels have led to few clinically useful insights in recent decades" (Eronen, 2019). Third, it may often be the *content* of psychological processes, rather than their physiological instantiations in the brain, that determine causality *at the psychological level* (Borsboom, Cramer, & Kalis, 2019). We know that patients who have severe fears of certain stimuli will avoid such stimuli, with all the emotional and behavioral consequences that go together with avoidance. And patients with delusions act according to the content of these delusions, information that helps in clinical care. It may not be necessary and potentially fully uninformative to study brain correlates of these psychological contents for determining their impact on other psychological constructs. Finally, the same contents, such as major depression or grandiose illusions, are likely multiply realizable, not only across the brains of different people but also in the brain of the same person over time (Borsboom et al., 2019).

While the levels problem remains a challenging one, I remain unconvinced that biological levels offer better explanatory power *per se* for the constructs we study. Similarly, I remain unconvinced that mechanistic explanations are the only way forward, and agree with Shapiro (2019) that "the more kinds of explanation available to scientists, the more explaining they will be able to do." Shapiro also provides a more detailed discussion on the merits of other kinds of explanations such as functional analysis, and challenges of mechanistic explanations (Shapiro, 2019).

Imprecise Constructs as Obstacles to Theory Formation

It is crucial to think carefully about what the target system is that our theory is meant to represent, and how the components of the target system can be measured so they precisely map onto the theory's structure. As Vaidyanathan (this issue) puts it: "strong theories may need strong constructs" (p. 334). She also reminds us that our discipline faces particular challenges for the question of what theories are *about*: many psychological constructs we study are subjective to some degree, complicating their measurement (Flake & Fried, 2020; Fried & Flake, 2018). I see two core challenges: construct definition and measurement.

Psychological constructs like depression are often not clearly defined in empirical research. Neumann recently reviewed a random draw of 100 empirical papers on depression published in clinical journals (Neumann, 2020); of these, only 6% provided definitions of depression. The rest contained either no information or merely gave phenomenological descriptions *about depression* (e.g., "depression is characterized by low mood and anhedonia"). But such descriptions are not definitions, in much the same way measles comes with a fever and Koplik's spots, but measles *is* not fever and Koplik's spots. This lack of clear definitions then further complicates the already challenging task of measuring people's internal states. Over 280 scales to measure depression have been used in the last 9 decades (Santor, Gregus, & Welch, 2006), and a recent review identified 19 different depression measures across 30 clinical depression trials (Mew et al., 2020). Many of these scales differ considerably in content (Fried, 2017), and arguably assess different constructs. When the majority of studies in a field define, operationalize, and measure a construct differently, but it is assumed everybody is talking about the same (kind of) thing, this poses a serious threat to scientific progress.

Formal models for psychological constructs may help with this measurement problem because they clearly describe both the theory's structure and the target system it represents (Robinaugh et al., 2019). Importantly, such models can be a valuable *starting point* for theory formation, because it is never too early to ask what components the target system has, how these relate to each other, and how they should be operationalized and measured (Robinaugh et al., 2020).

Theories Are About Real Entities

A final issue in the discussion of what theories are about is the question of whether psychological constructs are "real"

(and what this would mean), contrasting the notion that they are just summaries of data. Yarkoni (this issue) provides an instrumentalist point of view in his commentary that I endorse. My view on the matter, expressed for mental disorders below, also holds for other psychological constructs: "the most important feature of a diagnosis is not that it exists outside human classification systems as a real entity (Fine, 1984); above all, a diagnosis should be useful. And a diagnosis is useful if it provides clinical utility […]: it should help clinicians to determine prognosis, treatment plans, and potential treatment outcomes for their patients" (Fried, 2015).

I believe that psychological constructs are properties that emerge from complex systems, and whether the outcomes of such systems are considered real is not important to me. However, that does not change the fact that interventions on causal systems need to get causal aspects of a system right, and that, in turn, means whether the *p* factor causes psychopathology broadly, or provides a summary of psychopathology, is a crucial distinction irrespective of the question whether it exists outside of statistical models (Fried, Greene, & Eaton, 2020; Watts, Poore, & Waldman, 2019). Sidestepping discussions about the ontological status of constructs also does not mean that instrumentalists need not justify the selection of their statistical models, at least on statistical grounds: data can be summarized via different techniques, and reflective and formative latent variable models can lead to different outcomes (Rhemtulla, van Bork, & Borsboom, 2020; Widaman, 1993).

How to Develop Theories and Models

The target article does not provide much help on how to develop theories or models. In this section, I provide an overview of a recently proposed framework for theory formation (for other recent work on the topic, cf. Guest & Martin, 2020; Smaldino, 2020; Van Rooij & Baggio, 2020), and then critically discuss merits of verbal theories vs formalization that were brought up in several commentaries.

The Theory Construction Cycle

The *theory construction cycle* recently proposed by Borsboom et al. (2020) provides five steps to develop psychological theories. I will use the *p* factor to demonstrate the framework. The first step is the identification of a robust phenomenon, such as the fact that psychopathology symptoms are generally positively related in cross-sectional data. This positive manifold can be re-expressed as a latent variable referred to as the statistical *p* factor. Watts et al. (this issue) discuss whether this factor *should* be considered a robust phenomenon: it is a robust statistical effect, but its composition changes across datasets. This is not surprising, since the *p* factor, the result of several sophisticated reflective latent variable models, is nearly indistinguishable from the simple unweighted sum of variables that go into this model (Fried et al., 2020). And if *p* is not more than a summary of the data, it will differ when variables differ across datasets. Further, Watts et al. show that the most commonly used statistical technique to obtain the *p* factor (the bifactor model) overfits data while the literature has relied on model fit nearly exclusively in justifying the use of this model. For the sake of the argument, however, let us assume for now that we want to explain the presence of this statistical *p* factor.

The second step of the theory construction cycle involves the development of a proto-theory to explain the phenomenon. This proto-theory consists of a number of general principles expressed verbally. Several proto-theories have been proposed to explain the *p* factor. One is that *p* reflects a unitary, underlying causal construct for psychopathology broadly (Caspi & Moffitt, 2018; Krueger, 1999), another that *p* reflects impairment of functioning (because diagnoses of mental illness require the presence of impairment, people with more symptoms or diagnoses will naturally have higher levels of impairment). Watts et al. (this issue) argue that the most plausible proto-theories for *p* are not only the most flexible (i.e., least risky) ones—they can hardly be understood as explanatory principles. I would further argue proto-theories such as the idea of impairment (an index, not a cause) do not align with the reflective latent variable model the field has solely relied upon, and even strong proponents of the idea that such reflective models are causally agnostic grant that the models should not be used when there are no good reasons to believe that indicators share common causes (DeYoung & Krueger, this issue). Overall, formulating proto-theories may be a crucial opportunity for future progress in the field, and the *g* factor literature offers important lessons for *p* factor theory building (Conway et al., this issue).

To my knowledge, the three subsequent steps have not been carried out in the *p* factor literature, and I will summarize them only briefly. Third, verbal principles are translated into a formal model, which helps theorists think more clearly about the target system represented by the theory's structure, and enables simulating theory-implied data to find out what data would actually look like if the theory were true (Haslbeck et al., 2019; Robinaugh et al., 2020). Fourth, the adequacy of the formal model is explored: does it provide a sufficiently powerful explanation of the phenomenon under investigation? Does the theory-implied data look like real data, for example, does the model reproduce correlations among symptoms and diagnoses comparable to observed data? And fifth, theorists evaluate the theory's "overall worth" or "explanatory goodness." This includes criteria such as the explanatory breadth of the theory or simplicity. Borsboom et al. (2020) stress that the theory construction cycle will often run through numerous loops and provides but a rough framework; sometimes one may be able to skip steps, or go backward a step instead of forward. They also integrate their theory construction cycle with the well-known empirical cycle of empirical observation, formulation of hypotheses, development of hypotheses tests, testing hypotheses, and overall evaluation (De Groot, 1969).

The theory construction cycle reminds me of Vaidyanathan's (this issue) Goodrich quote: "theories look great on paper until reality scribbles all over the page." In my view, theoretical reforms should facilitate that our initial theories are inherently plausible and written down clearly, but also help us embrace the fact that reality will generally add in important ways to our theories: scribbling of reality is a feature, not a nuisance.

Benefits of Formalization

Given the strong ambiguity of many theories in psychological sciences that are usually expressed narratively, I agree with recent efforts to introduce psychologists to formalized theories. Van Rooij & Baggio (this issue) demonstrate that formalization can introduce important *choice points* in theory development, and Smaldino (this issue) describes formalization as *scrutinizing the systems* we are talking about.

Conway et al. (this issue) as well as DeYoung & Krueger (this issue) highlight the value of verbal theories. I agree in principle that ideas can be expressed precisely in a narrative way, addressing the problem of ambiguity. But there is a second core advantage to formalization: knowing what data are expected given a specific theory. A formal or computational model allows to simulate data under a theory, and enables the comparison of such theory-implied data with actual data; verbal theories lack this feature. We recently demonstrated this using the example of a very simple system (Robinaugh et al., 2020): a vicious cycle between two variables A $\rightleftarrows$ B, with an additional negative feedback loop on A to make sure A slowly decreases over time. Not only do small changes in the way the causal effects between A and B are implemented in the computational model (e.g., as linear or sigmoidal) dramatically change the theory-implied data—the data resulting from the model is unknowable without simulating data. And how can psychologists corroborate or update theories when they do not even know what sort of data their theory would produce? Formal models can be exceptionally helpful here and provide tools for thinking clearly, evaluating explanations, informing theory development, informing measurement, and facilitating collaboration and integration (Robinaugh et al., 2020).

Conclusion

How exactly can psychologists become better theorists and modelers? Conway et al. (this issue) propose that psychologists should receive better training in four areas: research techniques and methods, data diagnostics, philosophy of science, and theoretical psychology. Similarly, Smaldino (this issue) suggests that better training in statistical modeling and a firmer grasp of the philosophy of science will help overcome some of the problems the field faces. This is consistent with prior calls in the literature (Morton, 2009), and somewhat reminiscent of the statement Plato had famously engraved at the door of his Academy in Athens: "Let no one ignorant of geometry enter." Few people have made this point as forcefully as Meehl (1990b): "Inability to think mathematically among psychologists except in certain special areas is sometimes so gross as to be embarrassing to one familiar with the quantitative sophistication in other sciences."

Here is the problem: psychology is an inherently interdisciplinary field, and academic positions come with an overwhelming amount of expectations. Expected academic skills, expertise, and services that I collected from job postings, university promotion guidelines, grant guidelines, and workshops on career advancement in my discipline include (1) teaching and student supervision; (2) clinical expertise, up to the common expectation to be a researcher, teacher, and psychotherapist; (3) substantive expertise in the respective fields of interest; (4) empirical expertise, for example, how to conduct experiments in labs or run clinical trials; (5) measurement expertise, for example, how to collect task-based measures or neuroimaging data; (6) expertise in advanced data-analysis techniques; (7) and expertise in at least one programming language such as Python. In addition, like other academics, psychologists are being evaluated for (8) scientific service, such as reviewing papers and grants as well as editorial services for scientific journals; (9) university service, including participation in committees; (10) applying for grants; and (11) science communication. This list is far from complete (cf. Wright & Vanderford, 2017).

If we add even more rigorous training in math, modeling, and philosophy, I wonder if we'd be asking a little too much of psychologists. I see three solutions. First, we should consider removing some educational content to make space for the sort of training that Conway et al. (the issue) and Smaldino (this issue) propose. For example, I had to learn over 10 different basic emotion theories by heart in my studies (there are 4 basic emotions; no there are 5 basic emotions; no there are 8 basic emotions …), and keep wondering whether that was a good use of my time. Second, we should offer psychologists the possibility to become experts specifically in theoretical psychology, along with relevant mathematical and philosophical training. Theoretical biologists, theoretical physicists, and theoretical economists are usually not also experts in measurement and experimental research. And third, "remember that almost every topic tackled by researchers working in psychology departments is also worked on by researchers working in departments of anthropology, sociology, political science, economics, communication, biology, neuroscience, cognitive science, or computer science" (Smaldino, this issue, p. 300). Science is a team sport, and psychologists need not have all expertise required to conduct strong interdisciplinary work.

Speaking a common language across disciplines can help facilitate interdisciplinary research. Formal models provide such a language.

ORCID

Eiko I. Fried http://orcid.org/0000-0001-7469-594X

References

Apel, J. (2011). On the meaning and the epistemological relevance of the notion of a scientific phenomenon. *Synthese*, 182:23–28. www.jstor.org/stable/41477615

Bailer-Jones, D. (2009). *Scientific models in philosophy of science*. Pittsburgh, PA: University of Pittsburgh Press.

Barth, J., Munder, T., Gerger, H., Nüesch, E., Trelle, S., Znoj, H., … Cuijpers, P. (2013). Comparative efficacy of seven psychotherapeutic interventions for patients with depression: A network meta-analysis. *PLoS Medicine*, *10*(5), e1001454. doi:10.1371/journal.pmed.1001454

Borsboom, D., van der Maas, H. L. J., Dalege, J., Kievit, R., & Haig, B. (2020). *Theory Construction Methodology: A practical framework for theory formation in psychology*. Retrieved from https://psyarxiv.com/w5tp8/

Borsboom, D. (2013). *Theoretical amnesia*. Charlottesville, VA: Open Science Framework. http://osc.centerforopenscience.org/2013/11/20/theoretical-amnesia/

Borsboom, D., Cramer, A. O. J. J., & Kalis, A. (2019). Brain disorders? Not really: Why network structures block reductionism in psychopathology research. *Behavioral and Brain Sciences*, *42*, 1–63. doi:10.1017/S0140525X17002266

Borsboom, D., Fried, E. I., Epskamp, S., Waldorp, L. J., van Borkulo, C. D., van der Maas, H. L. J., & Cramer, A. O. J. (2017). False alarm? A comprehensive reanalysis of "Evidence that psychopathology symptom networks have limited replicability" by Forbes, Wright, Markon, and Krueger. *Journal of Abnormal Psychology*, *26*(7), 989–999.

Burger, J., Van Der Veen, D. C., Robinaugh, D. J., Quax, R., Riese, H., Schoevers, R. A., & Epskamp, S. (2020). Bridging the gap between complexity science and clinical practice by formalizing idiographic theories: A computational model of functional analysis. *BMC Medicine*, *18*(1), 1–18. doi:10.1186/s12916-020-01558-1

Caspi, A., & Moffitt, T. E. (2018). All for one and one for all: Mental disorders in one dimension. The American Journal of Psychiatry, *175*(9), 831–844. doi:10.1176/appi.ajp.2018.17121383

Cuijpers, P., Noma, H., Karyotaki, E., Vinkers, C. H., Cipriani, A., & Furukawa, T. A. (2020). A network meta-analysis of the effects of psychotherapies, pharmacotherapies and their combination in the treatment of adult depression. *World Psychiatry*, *19*(1), 92–107. doi:10.1002/wps.20701

Cummins, R. (2000). "How does it work?" versus "What are the laws?": Two conceptions of psychological explanation. In F. C. Keil & R. A. Wilson (Eds.), *Explanation and Cognition* (p. 117–144). Cambridge, MA: The MIT Press.

De Groot, A. D. (1969). *Methodology: Foundations of inference and research in the behavioral sciences*. Berlin, Germany: Mouton.

Eronen, M. I. (2019). The levels problem in psychopathology. *Psychological Medicine*, 1–7. doi:10.1017/s0033291719002514

Flake, J., & Fried, E. (2020). Measurement Schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*,

Fried, E. I. (2015). Problematic assumptions have slowed down depression research: Why symptoms, not syndromes are the way forward. *Frontiers in Psychology*, *6*(306), 309–311. doi:10.3389/fpsyg.2015.00309

Fried, E. I. (2017). The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *Journal of Affective Disorders*, *208*, 191–197. doi:10.1016/j.jad.2016.10.019

Fried, E. I., & Flake, J. K. (2018). *Measurement matters*. Retrieved from https://www.psychologicalscience.org/observer/measurement-matters

Fried, E. I., & Robinaugh, D. J. (2020). Systems all the way down: Embracing complexity in mental health research. *BMC Medicine*, *18*(1), 205–204. doi:10.1186/s12916-020-01668-w

Fried, E. I., Greene, A. L., & Eaton, N. R. (in press). The p factor is the sum of its parts, for now. *World Psychiatry*, 1–6.

Guest, O., & Martin, A. E. (2020). *How computational modeling can force theory building in psychological science*. Retrieved from doi:10.31234/osf.io/rybh9

Haslbeck, J., Ryan, O., Robinaugh, D., & Waldorp, L. (2019). Modeling *psychopathology*: From *data models to formal theories*. Retrieved from https://psyarxiv.com/jgm7f/

Hollon, S. D., & Beck, A. T. (2013). Cognitive and cognitive-behavioral therapies. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 393–442). Hoboken, NJ: Wiley.

Kendler, K. (2005). "A gene for.": the nature of gene action in psychiatric disorders. *The American Journal of Psychiatry*, *162*(7), 1243–1252. doi:10.1176/appi.ajp.162.7.1243

Krueger, R. F. (1999). The structure of common mental disorders. *Archives of General Psychiatry*, *56*(10), 921–926. doi:10.1001/archpsyc.56.10.921

Lewin, K. (1952). *Field theory in social science: Selected theoretical papers by Kurt Lewin*. London, UK: Tavistock.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: The slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, *46*(4), 806–834. doi:10.1037//0022-006X.46.4.806

Meehl, P. E. (1990a). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry*, *1*(2), 108–141. doi:10.1207/s15327965pli0102_1

Meehl, P. E. (1990b). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, *66*(1), 195–244. doi:10.2466/pr0.1990.66.1.195

Mew, E. J., Monsour, A., Saeed, L., Santos, L., Patel, S., Courtney, D. B., … Butcher, N. J. (2020). Systematic scoping review identifies heterogeneity in outcomes measured in adolescent depression clinical trials. *Journal of Clinical Epidemiology*, *126*, 71–79. doi:10.1016/j.jclinepi.2020.06.013

Morton, R. B. (2009). Formal modeling and empirical analysis in political science. In S. Pickel, G. Pickel, H.-J. Lauth, & D. Jahn (Eds.), Methoden Der Vergleichenden Politik-Und Sozialwissenschaft (pp. 27–35). Berlin, Germany: Springer.

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

Nettle, D. (2004). Evolutionary origins of depression: A review and reformulation. *Journal of Affective Disorders*, *81*(2), 91–102. doi:10.1016/j.jad.2003.08.009

Neumann, L. (2020). *Transparency in measurement: Reviewing 100 empirical papers using the Hamilton Depression Rating Scale*. Leiden, The Netherlands: Leiden University.

Peirce, C. S. (1931). *The collected papers of Charles Sanders Peirce* (Vols. 1–8). Cambridge, MA: Harvard University Press.

Rhemtulla, M., van Bork, R., & Borsboom, D. (2020). Worse than measurement error: Consequences of inappropriate latent variable measurement models. *Psychological Methods*, *25*(1), 30–45. doi:10.1037/met0000220

Robinaugh, D. J., Haslbeck, J. M. B., Ryan, O., Fried, E. I., & Waldorp, L. J. (in press). Invisible hands and fine calipers: A call to use formal theory as a toolkit for theory construction. *Perspectives on Psychological Science*, 1–11.

Robinaugh, D. J., Haslbeck, J. M. B., Waldorp, L. J., Kossakowski, J. J., Fried, E. I., Millner, A. J., … Borsboom, D. (2019). Advancing *the network theory of mental disorders*: A *computational model of panic disorder*. Retrieved from https://psyarxiv.com/km37w/

Santor, D. A., Gregus, M., & Welch, A. (2006). Eight decades of measurement in depression. Measurement: Interdisciplinary Research & Perspective, *4*(3), 135–155. doi:10.1207/s15366359mea0403_1

Scheel, A. M., Tiokhin, L., Isager, P. M., Lakens, D., Tiokhin, L., Scheel, A. M., & Isager, P. M. (2020). *Why hypothesis testers should spend less time testing hypotheses*. Retrieved from https://psyarxiv.com/vekpu

Shapiro, L. (2019). A tale of two explanatory styles in cognitive psychology. *Theory & Psychology*, *29*(5), 719–735. doi:10.1177/0959354319866921

Smaldino, P. (2020). How to translate a verbal theory into a formal model. Retrieved from https://econtent.hogrefe.com/doi/pdf/10.1027/1864-9335/a000425

Van Rooij, I., & Baggio, G. (2020). *Theory before the test: How to build high-verisimilitude explanatory theories in psychological sciences*. https://psyarxiv.com/7qbpr/

Watts, A. L., Poore, H. E., & Waldman, I. D. (2019). Riskier tests of the validity of the Bifactor Model of Psychopathology. *Clinical Psychological Science*, *7*(6), 1285–1303. doi:10.1177/2167702616673363

Widaman, K. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? *Multivariate Behavioral Research*, *28*(3), 263–311. doi:10.1207/s15327906mbr2803_1

Woodward, J. F. (2011). Data and phenomena: A restatement and defense. Synthese, *182*(1), 165–179. doi:10.1007/s11229-009-9618-5

Wright, C. B., & Vanderford, N. L. (2017). What faculty hiring committees want. *Nature Biotechnology*, *35*(9), 885–887. doi:10.1038/nbt.3962

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *12*(6), 1100–1122. doi:10.1177/1745691617693393