PSYCHOLOGICAL INQUIRY
2020, VOL. 31, NO. 4, 321–325
https://doi.org/10.1080/1047840X.2020.1853477

Routledge
Taylor & Francis Group

Theory Development Requires an Epistemological Sea Change

Iris van Rooij[a] and Giosuè Baggio[b]

[a]Donders Institute for Brain, Cognition and Behaviour, Centre for Cognition, Radboud University, Nijmegen, The Netherlands; [b]Norwegian University of Science and Technology, Trondheim, Norway

Introduction

Up until 2019, "psychological reform" mostly meant methodological and statistical reform of empirical research practices in psychology. Since then, however, we have seen a surge of proposals for theoretical reform (e.g., Guest & Martin, in press; Muthukrishna & Henrich, 2019; Smaldino, 2019; Szollosi & Donkin, 2019; van Rooij, 2019; van Rooij & Baggio, in press). While those calling for theoretical reform may agree on many things, they also do not form a monolith. For example, there are differences in emphasis placed on theories *vs* models, different views of what theories and models amount to and are supposed to achieve, and different opinions about which tools and concepts from other disciplines are most useful for building theories in psychology.[1] One aim of the present commentary is to highlight some of this diversity by commenting on Fried's target article in this broader context.

It is important that the pluralism that currently characterizes theoretical reform proposals is made visible to non-theoreticians. The risk of failing to do so is that potentially competing views and incompatible proposals may be unreflectively adopted and mixed in an *ad hoc* fashion, only leading to more problems. Theory reform has just started: we expect a development toward a better understanding of how theory can serve psychological inquiry; but to afford development we need exploration of ideas, diversity of views, and clear choice points—all of which need to be visible to both theoreticians and non-theoreticians alike.

While there are many points of agreement between Fried's and our own views, there are also points of divergence (more below). In short, we see Fried's target article as a valuable contribution to motivating an appeal to theory from within empirical research in psychology, without however providing a map of how to actually build and develop theory. Moreover, when Fried proceeds to assess what a good theory is, his focus seems to be mainly on good tests and empirical consequences of theory. In our opinion, if not supplemented with definite ideas about theory development, this approach risks stifling theoretical progress, as it fails to reveal the full potential of theory in psychological epistemology. It pitches theory in a way that may resonate well with empirical scientists, facilitating adoption of the *idea* of "theory," but without engaging in the necessary sea change in epistemology. In the remainder of this commentary, we explain our position.

We start by summarizing our interpretation of Fried's stance on the role of theory in psychology, and then highlight points of agreement. Next, we briefly summarize our own view on the role of theory in psychology, and then highlight points of disagreement. We close with a statement about the conceptual change that is needed for the field to start adopting and improving theoretical practices.

Our Reading of the Target Article

Fried's paper is a thorough discussion of the consequences of the lack of (explicit) theory in areas of psychology that make routine use of factor and network models. One thread running through the paper is the all-important contrast between statistical models and theoretical models: models of relations between measured or observed variables *vs* causal models of the underlying systems. For Fried, the primary aim of theory in psychology is to help "explain, predict, and control phenomena" (p. 271), not to provide (statistical) models of data. A good theory is formalized, explanatory, predictive, and theorist-independent; also, it should generate consequences for non-actual situations. A theory that lacks one or more of these virtues is more likely to lead to invalid inferences from data and less likely to create the conditions for theory to fail in informative ways, so as to be revised in the direction of greater verisimilitude. He shows (or rather reiterates a compelling argument) that factor and network models suffer from the problem of statistical equivalence, or equal fit to data, which makes it impossible to use such models to unambiguously and conclusively gain access to the underlying "data generating mechanism" (p. 274). It is precisely this failure of statistical models that, according to Fried, motivates the need for something more: theoretical models, or models of the causal systems that generate the observed data. The bulk of Fried's paper is a detailed analysis of the kinds of problems that can arise when no firm correspondence can be established between a statistical model and a theoretical model, if the latter is weak, absent, or

CONTACT Iris van Rooij i.vanrooij@me.com Donders Institute for Brain, Cognition, and Behaviour, Centre for Cognition, Radboud University, Montessorilaan 3, 6525 HR Nijmegen, The Netherlands.

[1] For instance, some may find inspiration in physics (Fried, this issue), some promote an evolutionary stance (Muthukrishna & Henrich, 2019), and others draw on principles from computer science (van Rooij & Baggio, in press).

just latent (causal beliefs that are not explicated). In what follows, we will focus on the epistemological assumptions of Fried's proposal, specifically his views on the structure and aims of theory, and how theory can be designed and constrained to achieve greater truthlikeness.

Points of Agreement

As noted, there is much we agree with Fried's general stance on the need for theory in psychological science, which is again shared with several other positions in current theoretical reform proposals. We believe that these areas of agreement can be boiled down to four general points.

What Theory Is Not

We agree with Fried that theory is not (not just, and not primarily) a set of statistical models that can be directly compared to empirical data. Statistical models may be able to capture relations between empirical variables with excellent fit (modulo statistical equivalence), but they do not, by design, specify the process that generates the behavior of the system, which may be partly (or even largely, in the case of psychology) unobservable. If statistical models fall short of providing access to the system's internal states and actions, so do verbal empirical hypotheses about effects, of the form "variable X should show larger/smaller measured values than variable Y" or "variables X and Y should show a positive/negative linear correlation," etc. None of this is "theory," and generally speaking no amount of formalization or technical refinement can turn a statistical model (of data) into a theoretical model (of data generating mechanisms) that can (a) address the damning problem of equal fit to data and (b) provide explanations of genuine psychological phenomena (such as mental capacities, see below) as opposed to merely capturing patterns in data. So, psychological inquiry needs more than just statistical models: but what then? We return to this in a moment.

Theory Needs Formalization

We agree that theory requires formalization. Fried does not delve much into this aspect of theory development, but his suggestions all seem on the right track. For example, he sees (and we agree) formalization as a way of making theory precise, less ambiguous (because less dependent on natural language and on the theorist's own interpretation and inferences), and easier to test. We also agree that the same verbal theory can be formalized in several different ways, and that different formalizations can lead to effectively different theories, with different predictions, etc. Formalization introduces important choice points in the process of theory development.

Most Theory in Psychology Is Weak

We share Fried's diagnosis that most current theories in psychology are "weak": "narrative and imprecise accounts of hypotheses, vulnerable to hidden assumptions and other unknown" (p. 272). Clearly, not all psychological theories are like that, and weakness may be just a property of a theory in its early stages of development, rather than an intrinsic limitation. But it is fair to say that there are general challenges, experienced by every theorist and ever theory, involved in developing precise, explicit theories and models of psychological phenomena: formalization is just one such challenge; imposing the right constraints on formal theory is another (Fried touches on this issue in section 5.4). The open question is how to actually proceed to strengthen theory.

Poor inference from lack of theory

Finally, we agree that weak, absent, or latent theories diminish our capacity to make valid inferences to and from data (e.g., about how to explain data, what follows from theory, etc.). Current psychological science, and its associated epistemology, have largely focused on just one type of uncertainty that potentially mars the robustness of inferences to and from data: statistical uncertainty, e.g., as linked to probabilities of accepting or rejecting hypotheses about effects in a given population, based on data from a random sample. But, absent strong theory, the effects of this uncertainty are compounded by the logical uncertainty intrinsic to drawing inferences from incomplete or unwarranted premises: unexplicated assumptions about theory or models, unclear distinctions between statistical and theoretical models etc.

Intermezzo: Our View of Theory

Before we move on to points of disagreement with Fried's position, we briefly summarize our view of theory and of the role of theory in psychological inquiry. For more details, we refer to (van Rooij & Baggio, 2020).

Our view of theory is rooted in the philosophy of psychological explanation (e.g., Bechtel & Shagrir, 2015; Cummins, 1985, 2000; Egan, 2010, 2017; Wright & Bechtel, 2007). It differs in crucial ways from traditional philosophy of science, which has been dominated by the philosophy of physics. The main aim of theories, as we conceive of them, is to provide *explanations* of key phenomena that collectively define the field of study of psychology (as opposed to, say, chemistry, biology or sociology). Data and effects derived from measurement or observation and statistical inference are only secondary explananda for psychology. The primary explananda are *capacities*. Relevant capacities for psychology may span different levels of organization of complex systems: e.g., cognitive capacities (language understanding, reasoning, decision-making, etc.), capacities for social interaction (coordination, competition, communication) and cultural evolution (transmission and acquisition of language and social norms), but also the capacities of neurons (spiking, firing) or neuronal interactions (exciting, inhibiting) that presumably implement psychological processes in biological brains. A capacity can be understood as a more or less reliable ability (or disposition or tendency) to transform some initial state into a resulting state. The resulting states need not be "desirable," and the tendency of a person, given certain initial conditions, to more or less reliably converge on a particular state of mind is a capacity, too. In this light, depression and anxiety (but also mental well-being) can theoretically be seen as states to which our mental states converge under certain conditions, but not others.

How do theories explain capacities? Psychological explanations are a species of *mechanistic explanation*: they aim to explain target phenomena (explananda) by postulating component processes and specifying how their operation and interaction give rise to, i.e., *produce*, the target phenomenon. One influential form of mechanistic explanation is *computational explanation*, which casts mechanisms in computational terms and using tools from computer science and related areas of the formal sciences. Our view of theory builds on the paradigm of computational mechanistic explanation. It distinguishes different levels of explanation (Marr, 1982) and adopts a top-down approach in the process of building theories, starting at the top (computational level) and working downwards to lower levels (algorithmic and implementational levels). Capacities are characterized at the level of the formal theory, and algorithmic and implementational level explanations characterize the processes and physical realization, respectively.

Points of Disagreement

In this section, we frame the issues on which we diverge the most from Fried's proposal in a way that serves best the aim of getting our own points across. We hope our earlier summary and points of agreement with Fried suffice to guard against any potential sense of misrepresentation. Subheadings below refer to our own positions. We welcome clarifications where we may have misunderstood Fried's views.

Effects Do Not Provide "Solid Stones" for Theory

Fried opens his article by quoting (Muthukrishna & Henrich, 2019):

> The present methodological and statistical solutions to the replication crisis will only help ensure solid stones; they don't help us build the house.

This is an apt metaphor for how methodological and statistical reformers conceive of the relation between data and theory: methodological and statistical solutions (such as "preregistration") make sure that we at least first get the "effects" right. As Fried puts it:

> In the best case, these new best practices lead to more reliable and replicable statistical effects, i.e. robust phenomena (…) usually the relationship between two variables, or the difference of two groups (…) (p. 271)

These robust and reliable effects can then provide the solid stones from which to build theory.

In our view, this metaphor however risks misconceiving the basis of theory, both its materials and its starting point. 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.[2] We previously used a different analogy to make this point (van Rooij & Baggio, in press):

> (…) trying to build theories on collections of effects is much like trying to write novels by collecting sentences from randomly generated letter strings (…) the majority of the (infinite possible) effects are irrelevant for the aims of theory building, just as the majority of (infinite possible) sentences are irrelevant for writing a novel. Also, many of the relevant effects (sentences) we may never happen upon by chance, given the vast space of possibilities.

What is needed is a different starting point for building theories: psychology's primary explananda are capacities, not effects (see also the section "Intermezzo: Our View of Theory" and Cummins, 2000). Capacities are real-world phenomena, not to be confused with the kind of effects (secondary explananda) that can be statistically established in our labs. By building good candidate theories of psychological capacities we can have solid ground for inquiry into effects. Effects are usually not the stones with which we build theories, but sometimes they can be the windows that provide us the view of the surroundings of our house so that we can decide whether or not we want to keep living here. If our scenic observations generate too many errors of prediction and explanation (and control, Fried would add), then we may decide to close shop and build a new house elsewhere.

Our view is not new. The idea that we can build a theory on theory-neutral "facts" has been philosophically discredited long ago, although discussions on several details continue: all interpretation of data is necessarily theory-laden (Hanson, 1958; Kuhn, 1962; Shapere, 1982). The only reason we sometimes forget that is that our theoretical assumptions lay implicit and unexamined. Fried is, of course, aware of this fact; but it is unclear to us why then he appears to hold on to the idea that effects form the basis on which we build theory.

Theories Provide Mechanistic Explanations

A second form of disagreement appears to be what counts as "explanation." Fried seems to want theories to explain (besides predict and control) phenomena (e.g., "Strong theories provide a clear explanation of a phenomenon"). Yet, Fried says relatively little about what it takes for something to be an explanation of a phenomenon. What we could distill is that Fried seems to subscribe to an axiomatic view of theories, and presumably (a fortiori) to a deductive view of explanation, as he writes:

> I understand theories as sets of axioms or assumptions that help explain, predict, and control phenomena. (p. 271)

and later, again:

[2] And when inquiry into effects is guided by theory and sound scientific and statistical inference (Devezer et al., 2020; Navarro, 2019), methodological guidelines for preregistations seem redundant, at best (Szollosi et al., 2020).

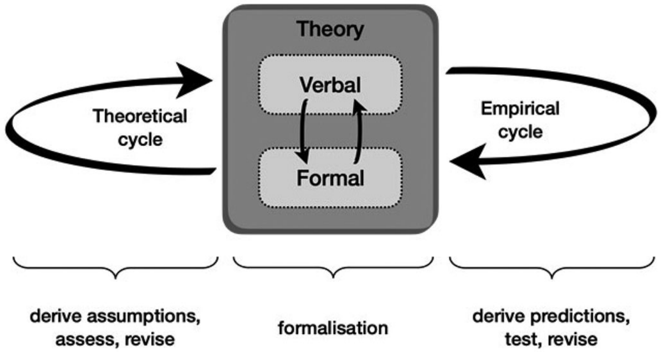
Theory
Verbal
Formal
Theoretical cycle
Empirical cycle
derive assumptions, assess, revise
formalisation
derive predictions, test, revise

Figure 1. Theory development requires the interplay of verbal and formal analyses of a capacity, a theoretical cycle aimed at endowing the theory with greater a priori plausibility (verisimilitude), and an empirical cycle aimed at assessing the theory's empirical adequacy. Testing theories in the empirical cycle can be guided by first selecting theories with greater verisimilitude as may result from the theoretical cycle. Adapted from van Rooij and Baggio (2020).

> Strong theories (…) explicate a precise set of assumptions and axioms about a phenomenon unambiguously.

The "strength" that Fried attributes to this notion of theory is that it should allow for unambiguous formalization and theorist-independent predictions. Yet, since neither formalization nor prediction are sufficient to yield explanation, it is not clear what is Fried's notion of explanation, or whether he actually equates explanation with prediction, or at best, underemphasizes explanation in favor of a more central role for prediction. The latter would fit with his emphasis on theory testing in a paper about the role of theory in psychology (more below).

An axiomatic approach may make sense if one is studying the principles of fundamental physics, in search of laws that apply to all systems, always and anywhere, and in search for explanations that are effectively instances of subsumption of phenomena under those laws. However, an axiomatic approach seems unsuitable for special sciences, such as psychology, where we seek explanations of why and how certain types of systems (e.g., neurons, brains, people, groups) have the capacities that they have, under "normal conditions." Explanations of capacities are mechanistic explanations: they need to be cast in terms of component processes and make clear how their organization and interaction produce those capacities (Cummins, 2000).[3] Psychological theories should therefore include representations of relevant mechanisms, framed in terms of mathematical models of qualitative structure (e.g., internal states, architectures, algorithms, etc.). These models of qualitative structure are not deduced from axioms or general laws, but are typically the result of a constructive process, which is invoked precisely in response to the fact that mental structure is unobservable—ontologically unobservable, that is, and not simply unobservable given the limitations of measurement instruments or of our own sensory apparatus. Fried gestures in the direction of qualitative structure when he mentions "data generating mechanisms," but since we submit, and Fried agrees, that phenomena are not the same as "data," and Fried thinks the explanantia are effects that are hidden in data + noise (p. 4), and are not capacities, we have difficulty interpreting this as a notion of mechanistic explanation of capacities.

Good Theories Produce *A Priori* Plausible Explanations

Fried seems to subscribe to the view that theories, in order to be "good" or "strong," need first and foremost make testable predictions. We think this is a widespread view in psychology. However, we disagree. Fried's position emerges quite clearly when he writes:

> (…) explicated theories are often weak theories: imprecise descriptions vulnerable to hidden assumptions and unknowns. Such theories do not offer precise predictions, and it is often unclear whether statistical effects actually corroborate weak theories or not. (p. 271)

While we do agree that theories are (and should be) tested, we think theories are neither *for* testing, nor *for* prediction. Instead theories are *for* understanding through explanation. Accordingly, we believe theories are "good" or 'strong', if they provide plausible explanations of capacities.

Testing is a means to an end, not a goal in itself. 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 (see disagreement 1). Although the distinction between weak theories and weak tests should be obvious, we find in Fried's paper no clear distinction between what makes a good *theory* and what makes a good *test* of theory. But as Cummins (2000) notes:

> [a] way in which talk of explanation in the context of the statistical analysis of data is likely to be misleading is that, even though experimenters sometimes are attempting to test a theory (…), this is an exercise in confirmation, not explanation.

So if good tests (or testability) do not define "good theory," and the ability to make precise predictions does not make for "good theory," what does? As said, theories in our view provide mechanistic explanations of capacities. Such theories are "good" or "strong" insofar as they have (at least) *a priori* verisimilitude. Theories that meet this criterion can be constructed in two general phases (see Figure 1):

1. **Formalization:** Informal ideas about the mechanism that produces the capacity (verbal theory) that have been abduced from observation and background knowledge are formalized (formal theory);
2. **Theoretical cycle:** The formalized theory is analyzed for its *a priori* plausibility (using formal tools and relevant background knowledge), and the theory is updated, revised and refined accordingly.[4]

[3] We see no theoretical reason why such explanations cannot also be had for the types of phenomena that Fried is interested in (anxiety, depression, etc.).

[4] The theoretical cycle is a generalization of what van Rooij et al. (2019) called the tractability cycle, where the theoretical tests are of one particular type. In the theoretical cycle, any relevant theoretical tests may be performed, e.g., evolvability, learnability, developability, emergeability, physical realizability (see van Rooij and Baggio (in press) and van Rooij and Blokpoel (2020) for more details).

Here, phase 1 is aimed at ensuring that the theory is genuinely explanatory and well-defined, and phase 2 is aimed at endowing the (revised) theory with greater *a priori* plausibility (verisimilitude) prior to empirical testing, and further revising and refining the theory via the empirical cycle. While this picture gives an idealized view of the scientific process, in reality (actual research practice) the integration of the cycles will be more messy and not strictly sequential. Our point is not that theory is strictly temporarily prior to *all* empirical testing, but that it is epistemologically prior.

Conclusion

Reading the target article, it is difficult to resist the conclusion that Fried sees theory's primary role as a way to improve statistical inference and primarily address problems that arise at the interface between statistical models and data. While this provides motivation for empirical scientists to look at and appreciate the idea of (formal) theory, the epistemological pull toward theory does not appear to be powerful and attractive enough from Fried's perspective. Interest in theory development is likely to be stopped in its track when theory is pitched solely as something that improves psychologists day-to-day empirical work (empirical practices): experiments and statistical testing, with an eye to the distant aim of mechanistic explanation. For psychologists to start seeing the full potential of formal and cumulative theory development, and for them to start investing the necessary resources in it (theoretical practices), they (we) will need to let go of entrenched ideas that theories are built on effects, that theories are for testing, and that good theories are ones that make precise and testable predictions. Psychological science requires an epistemological sea change, a conceptual shift that allows us to appreciate that theories should provide formalized, mechanistic explanations of capacities, that theories are good when they do that in an *a priori* plausible manner, and that tests only serve to refine theoretical possibilities that we are already entertaining.

ORCID

Iris van Rooij http://orcid.org/0000-0001-6520-4635
Giosuè Baggio http://orcid.org/0000-0001-5086-0365

References

Bechtel, W., & Shagrir, O. (2015). The non-redundant contributions of Marr's three levels of analysis for explaining information-processing mechanisms. *Topics in Cognitive Science*, *7*(2), 312–322. doi:10.1111/tops.12141

Cummins, R. (1985). *The nature of psychological explanation*. Cambridge, MA: MIT Press.

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

Devezer, B., Navarro, D. J., Vandekerckhove, J., & Buzbas, E. O. (2020). The case for formal methodology in scientific reform. *bioRxiv*, 10.1101/2020.04.26.048306.

Egan, F. (2010). Computational models: A modest role for content. *Studies in History and Philosophy of Science Part A*, *41*(3), 253–259. doi:10.1016/j.shpsa.2010.07.009

Egan, F. (2017). Function-theoretic explanation and the search for neural mechanisms. In D. M. Kaplan (Ed.), *Explanation and integration in mind and brain science*. Oxford: Oxford University Press.

Guest, O., & Martin, A. E. (in press). How computational modeling can force theory building in psychological science. *Perspectives on Psychological Science*.

Hanson, N. R. (1958). *Patterns of discovery*. Cambridge: Cambridge University Press.

Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: Freeman.

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

Navarro, D. J. (2019). Between the devil and the deep blue sea: Tensions between scientific judgement and statistical model selection. *Computational Brain & Behavior*, *2*(1), 28–34.

Shapere, D. (1982). The concept of observation in science and philosophy. *Philosophy of Science*, *49*(4), 485–525. doi:10.1086/289075

Smaldino, P. (2019). Better methods can't make up for mediocre theory. Nature, *575*(7781), 9 doi:10.1038/d41586-019-03350-5

Szollosi, A., & Donkin, C. (2019). Arrested theory development: The misguided distinction between exploratory and confirmatory research. *PsyArXiv*, 10.31234/osf.io/suzej.

Szollosi, A., Kellen, D., Navarro, D. J., Shiffrin, R., van Rooij, I., Van Zandt, T., & Donkin, C. (2020). Is preregistration worthwhile? *Trends in Cognitive Sciences*, *24*(2), 94–95. doi:10.1016/j.tics.2019.11.009

van Rooij, I. (2019). *Psychological science needs theory development before preregistration*. Psychonomic Society. https://featuredcontent.psychonomic.org/psychological-science-needs-theory-development-before-preregistration/.

van Rooij, I., & Baggio, G. (in press). Theory before the test: How to build high-verisimilitude explanatory theories in psychological science. *Perspectives on Psychological Science*.

van Rooij, I., & Blokpoel, M. (2020). Formalizing verbal theories: A tutorial by dialogue. *Social Psychology*, *51*(5), 285–298. https://doi.org/10.1027/1864-9335/a000428.

van Rooij, I., Blokpoel, M., Kwisthout, J., & Wareham, T. (2019). *Cognition and intractability: A guide to classical and parameterized complexity analysis*. Cambridge: Cambridge University Press.

Wright, C., & Bechtel, W. (2007). Mechanisms and psychological explanation. In P. Thagard, D. Gabbay, & J. Woods (Eds.), *Philosophy of psychology and cognitive science*. Amsterdam, Netherlands, North Holland: Elsevier.