

Commentary

CBT for Public Good: Why We Need to Be More Comfortable Using Someone Else's Toothbrush

Laura D. Seligman

University of Texas Rio Grande Valley

The promise of behavior therapy and, later, cognitive behavioral therapy, was to alleviate human suffering by using a scientific approach to better understand how to develop, test, disseminate, and implement sound psychological treatments. As a field, we have seen great success, including the development of an ever-growing list of evidence-based treatments. However, the proliferation of “new” treatments does not seem to be matched by a commensurate growth in the number of people we are effectively able to help. Suggestions for how to address our stalled progress have included intensified efforts at dissemination and a focus on open science and replication. Here I suggest that if we are to make meaningful progress, the type of progress that makes a difference in people's lives, we must focus more on the why behind what we do—our theories. In particular, I argue that we must focus on formalized theory, but that for this type of work to flourish, we need to redefine and reshape the contingencies in the field of cognitive behavioral science, rethinking our ideas of how success should be defined.

Keywords: theory; cognitive behavioral therapy; philosophy of science

The author wishes to thank the anonymous reviewers for their suggestions that contributed to the final version of this paper.

The author declares no conflicts of interest.

This paper is based on the Presidential Address delivered by the author at the 56th Annual Convention of the Association for Behavioral and Cognitive Therapies, New York, NY.

Address correspondence to Laura D. Seligman, Ph.D., Department of Psychological Science, University of Texas Rio Grande Valley, 1201 W. University Drive, Edinburg, TX 78539. e-mail: laura.seligman@utrgv.edu.

0005-7894/© 2025 Association for Behavioral and Cognitive Therapies. Published by Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

AS WE GATHERED for the 56th Annual ABCT Convention, it marked the first time in 2 years that we did so in person. I believe it is not hyperbole to say that we had all been profoundly changed since seeing each other last. Some of us lost friends, family members, or coworkers. Almost all of us experienced significant disruptions in how we work and how we connect with family and friends. Even those of us fortunate to live in the wealthiest countries in the world, used to enjoying the buffer this privilege affords us, experienced an awareness of our vulnerability that could only be matched, perhaps, by one other event in recent history. There is no doubt that we have had many dark days these past few years but, as I believe is often the case in times of great upheaval, our search for meaning can create silver linings. For me, and my observations of friends and colleagues suggest that I am not alone here, these events have caused me to take time to pause and reflect on my priorities and values, both personal and professional, and to assess whether the way I live and work are consistent with these priorities.

I began my career with confidence that collectively, we—those of us who practice, teach, and study cognitive behavioral therapy (CBT)—would make slow but steady and meaningful progress, building a cumulative science that would be used to refine CBT in ways that had a positive impact on our communities. I fully expected that the CBT I practice and teach today would have evolved significantly from the CBT I was first introduced to in graduate school. But, as I watched fellow scientists react with precision and an almost unbelievable swiftness to address the novel coronavirus that had affected the world, and even as it became clear that the solution to this pandemic lay as much in human behavior as virology and

immunology, I did not see evidence of a precise and sophisticated science of behavior change being harnessed to inform public policy. There are many possible explanations for why this did not happen, some of which have very little to do with the progress, or lack thereof, of our science. Nonetheless, at some point I felt forced to question my early assumptions that, although incremental, our work was building toward something greater than the sum of its parts, that we were producing the type of work that brings about meaningful change. This is not the first time this question has been asked at ABCT. In fact, as we celebrated our 30th birthday in 1996, two special issues of *Behavior Therapy* were devoted to the question of whether behavior therapy was making progress in fulfilling its early promise (Forsyth & Hawkins, 1997). Now as we inch closer to our 60th birthday, it is notable that I am still asking the same question: Collectively, is our behavior consistent with ABCT's mission to alleviate human suffering?

If our attention is elsewhere, we must be clear about what is at stake. The pain, suffering, and lethality of mental illness is something we do not often discuss openly, except in the most dispassionate terms. There are good reasons for this, but we must not let the implications of our choices, for those in whose service we work, fade from our consciousness. In 2021, suicide was the 11th leading cause of death in the United States (Centers for Disease Control, 2021). This translates into over 48,000 Americans who lost their lives. Putting that number in perspective, even in the height of the COVID-19 pandemic, Americans between the ages of 10 and 34 were more likely to die by suicide than by COVID-19. In 2022, close to 107,491 Americans died due to overdose (Spencer et al., 2024). Approximately 10,000 Americans die from eating disorders each year (Deloitte Access Economics, 2020). These numbers are not exhaustive; they only begin to capture the lives lost and they do nothing to describe the suffering that does not result in death. Imagine going to your local emergency room in agony due to a broken leg and being sent home with some encouraging words and a referral because you are not at risk for imminent death, then returning home to find out that there are no appointments available to you for months. This would be unimaginable, but this is not unlike the situation faced by many who visit emergency rooms due to anxiety, depression, substance abuse, or a myriad of other conditions that get cordoned off as mental health.

It is our job to change this: whether we do this through teaching, research, or clinical service, all

of our professional activities must be centered around this goal. Ethically, I see no other conclusion to be possible. But how much of our daily efforts are focused on making meaningful change for those who should be at the center of our work? How successful have we been? The early promise of CBT was twofold: those that came to us would get treatment that had withstood scientific scrutiny, and CBT would continue to improve as the result of scientific progress. Are we on the path to realizing that promise? For those in the depths of depression, substance use, and other mental health crises, the pain is agonizing and real. The data show that still, many of these individuals will never get treatment that is grounded in science (e.g., Cooper & Kelland, 2015; Cox & Codd, 2024; Wolitzky-Taylor et al., 2015). For the few who do, this treatment might not look substantively different than it did 20 or 30 years ago, a fact underscored by the finding that the effect sizes for CBTs for anxiety disorders, arguably one of our areas of greatest success, have not significantly improved over the last 30 years (Hofmann et al., 2025). I would argue that we can and *must* do better but that the contingencies in our field are not only distracting from what should be our primary goal—to alleviate the suffering of those in our communities—but that these contingencies may even be antithetical to this goal. To find our way forward, here I will argue that we must invest in a renewed focus on theory-driven cognitive behavioral therapy, and I will contrast this with what I see in our field today. Further, I believe that if we are to accomplish the goal I set out here, our field must make fundamental changes in what we value, and the behaviors we choose to reinforce.

Our History With Theory

As I engaged in discussions at the highest levels of ABCT, I was surprised to learn that for a field that values clear operational definitions when discussing others' behavior, there was little consensus on the definition of CBT—on what CBT is and, even more pointedly, what CBT is not. ABCT's website states

Cognitive Behavioral Therapy is a type of treatment that is based firmly on research findings. It places emphasis on changing your cognitions (thoughts) or behaviors (actions) in order to effect change in how you feel. These approaches help people in achieving specific changes or goals.¹

¹ Note there is no reference on the ABCT website to indicate whether this is a definition that has been formally approved and endorsed by the board or the membership.

Without any reference to the mechanisms of change, definitions like this are broad enough to include activities that members of ABCT would certainly not consider under the umbrella of CBT, even activities like brainwashing and torture could arguably be defined as CBT using this definition. Even if we were to include the notion that CBT is limited to methods used to improve mental or physical health, one could argue that medications that reduce rumination or increase positive activity scheduling would be forms of CBT.

On the other hand, at times the definition of CBT has been interpreted narrowly enough that treatments that seem to fit squarely within the definition of CBT are now sometimes considered something related to CBT but not quite CBT (Hofmann et al., 2010). I would argue that the problem we face has is that our efforts in defining and advancing CBT have often been focused on *what* we do, the activities involved and how much they look the same or different than what has come before, and not enough on *why* we do it—the underlying theory that unites CBT.

Given our history, I recognize this assertion is not without controversy. CBT (or rather, one of its precursors, behavior therapy) did not arise from a unified organized movement but rather it seems to have developed organically as ideas from thought leaders and research findings blended into something that was clearly different than the approaches to mental health treatment that prevailed at the time. Naturally, the movement seems to have coalesced around common ground. Perhaps, equally unsurprisingly, it appears much of what was agreed upon was not so much what behavior therapy was but what it was not, owing to the dissatisfaction with the psychoanalytic model. Seemingly as a result, many early definitions of behavior therapy emphasized how it differed from psychoanalytic approaches (e.g., Eysenck, 1959). To be clear, for some, the focus on empirical science to affect behavior equated to learning theory and some would argue that behavior therapy is the application of learning theory to improve health and well-being (for an example of this perspective, see Kazdin, 1978), with CBT being an extension of this work. However, for a significant number, behavior therapy is defined primarily by its focus on the observable along with the adherence to science to inform practice (Goldfried & Davison, 1994) while remaining agnostic about the mechanisms used to achieve behavior change. From my observations, it appears we have agreed to disagree, and this has served us well to some extent—we have built a large tent, expanding the reach of CBT.

However, it seems we are almost at the point that our efforts to increase inclusivity, to ignore the differences in our thinking, have brought us to the point that CBT risks becoming a “brand” with little intrinsic meaning. We risk allowing our goal to become expanding this brand, with less of a focus on quality control or product improvement. With this perspective our priorities become self-focused, how to market our brand to achieve a bigger market share, and the goal to increase access to evidence-based treatments that continuously improve as a result of the cycle of scientific inquiry becomes secondary. However, we have known for some time focusing on the former (more access) without a focus on the latter (evidence-based treatments) is most certainly *not* better (Bickman, 1996; Weisz et al., 1997), at least not for those who receive these treatments and those who pay for them.

I believe to turn the tide here and prevent us from going down this path, we must focus on a more difficult question—at its core, What is CBT?—that a focus on the theory that underlies CBT is key to moving our field forward.

Building a Cumulative Science: Theory and Toothbrushes

To make the argument for theory, we must first agree on a basic definition of what is meant by this term. This is a difficult proposition as there is no universally accepted definition. To advance a shared understanding here, I adopt a basic definition of theory, borrowing from Fried (2020b), as broad but precise explanations of phenomena of interest, broad, in that theories are explanations of real-world phenomena in all their complexities but precise in that good theory is detailed enough to allow for falsification.

Why do I believe that theory-driven CBT is critical for progress? The answer is quite simple. I believe the alternative is to pursue a path doomed to failure if only because of its sheer impracticability. Decades of disconnected research findings that are not integrated into broader explanations will not solve real-world problems. This is not a new observation. In fact, the lack of theory in psychological science and the ramifications has been a topic of discussion since at least the 1970’s when Allen Newell highlighted the exact issues with which we continue to struggle today in an essay entitled “*You Can’t Play 20 Questions With Nature and Win*” (Newell, 1973). Here, Newell explains what he saw as the prevailing approach we have taken to moving forward the science of psychology:

The proper tactic is to frame a general question, hopefully binary, that can be attacked experimentally. Having settled that bits-worth, one can proceed to the next. The policy appears optimal—one never risks much, there is feedback from nature at every step, and progress is inevitable. Unfortunately, the questions never seem to be really answered, the strategy does not seem to work. (Newell, 1973, p. 8)

Newell goes on to lament the fact that the findings from our individual studies are never integrated into a more meaningful whole, and, in fact, he suggests that the way we do our work makes this goal virtually impossible. Newell's remarks were made specifically in reference to the field of cognitive psychology approximately 50 years ago, but I believe they aptly apply to the state of our knowledge of CBT as it exists today. For I believe, we have taken a strength of CBT—an empirical approach to the alleviation of human suffering—and interpreted this so narrowly, as to allow it to become our downfall.

How We Play Today's Game of CBT

In practice, this focus on empiricism has evolved into an almost exclusive focus on null hypothesis significance testing (NHST). To be clear, when referring to NHST, I mean not only the statistical tests used but also the general approach and process we use to test our hypotheses, as well as the practices and traditions that stem from our overreliance on NHST. So, the issue is not NHST testing itself but the culture that naturally follows from how successful science is defined when NHST is our primary tool.² NHST starts with a hypothesis offered by the researcher, ideally one that is grounded in past research or theory, but even a cursory review of many of the journals in our field reveals this is frequently not the case (e.g., McPhetres et al., 2021). In fact, often the justification for a hypothesis may simply be its novelty (i.e., the “no one has looked at this before” justification). This is not accidental, as novelty is a core currency in NHST—to add something of value in this system one must be asking new questions. These hypotheses typically take the form of treatment A is better than treatment B or group A will

score higher than group B on measure X or variations in which these relationships are framed in terms of covariation. Undoubtedly, almost every ABCT member will have at least some familiarity with this process, even if it is just a distant memory from graduate school days. However, implicit in the NHST approach is that every study is built on a multitude of auxiliary hypotheses (Fried, 2020a), for example, the sample used to test one's hypothesis, the measures used to assess the relevant constructs, and the setting in which the study is conducted. So, although we may be interested in testing whether Treatment A is superior to Treatment B, in practice we test whether Treatment A is better than Treatment B, in a sample of Caucasian middle-class adults between 18 and 40 years old, presenting to a university clinic, selected using a specific diagnostic interview, with outcomes measured using a specific measure, and the treatment delivered by a specific type of treatment provider.

This has several important implications. First, even in the most fortunate case in which our findings support our hypotheses, what we have learned is quite narrow; we have found support that Treatment A is superior to Treatment B under the narrow conditions created by our auxiliary hypotheses. That is not to say that Treatment A is not generally superior to Treatment B, in other words, this does not necessarily mean that our findings would not generalize, but that we simply do not know. If we want to know if Treatment A is generally better than Treatment B, we must ask this question repeatedly, with slight changes in our auxiliary hypotheses.³ Second, data inconsistent with study hypotheses seldomly result in a dismissal of the hypothesis, both because of the inability to prove the null and because often auxiliary hypotheses are invoked to maintain the plausibility, or at least the possibility, of the central hypotheses (Fried, 2020a). Moreover, it is also entirely possible that, given the logic underlying NHST, null findings are never communicated to the scientific community (i.e., the file drawer problem). So, even though one could argue that many studies could accomplish what a single study could not—that is, if a given hypothesis was tested in the context of several different sets of auxiliary hypotheses seldomly receiving support, the result

² Although it is outside the scope of the current paper to fully discuss how NHST contributes to our current scientific culture and how, in turn, that culture then interacts with the weaknesses inherent in NHST, the interested reader is referred to Lash (2017) and Meehl (1967, 1990).

³ This is true at least when we continue to operate under our current system. Likely, some variations in auxiliary hypotheses are trivial enough that generalizability could be assumed without the proof from an independent investigation but without theory to guide our thinking about the boundary conditions for an effect, we have little to guide us (Haefel et al., 2009).

would be dismissal of said hypothesis—it is highly likely this accumulation of null findings would never come to light. Importantly, under these conditions, we do not ever disprove anything, *despite this being a primary goal of science* (Haeffel, 2022). So, we have set up for ourselves a system in which one must find *something*; failure to support one's hypotheses means the inability to publish. In this way, we have built a system in which our personal goals, proving our hypotheses to be correct, conflict with the broader goals of science, identifying where we are wrong to correct our theories.

So, in sum, in the game of NHST, success means testing many specific, narrow, and new hypotheses, and supporting your predictions. High productivity, finding what you said you would find, and high novelty are what gets reinforced in such a system. Today, this is even more true as cognitive behavioral researchers judge ourselves and our colleagues on a variety of indices that we infer to equate to impact. As behaviorists whose field was once centered on the idea that one's environment shapes one's behavior, we must ask what types of behavior we are eliciting in a system that rewards high novelty and high productivity as well as always being right. Fortunately, the scientific literature has some answers for us.

Smaldino and McElreath (2016) examined this very question, using computational modeling to examine the natural selection of scientific labs whose continued existence and ability to create progeny (i.e., trainees to carry out the lab's work and traditions) are determined by the incentives with which we currently work. Simply put, they found two major trends emerge. First, effort decreases over time. Importantly, here effort equates to methodological rigor. Essentially, in a system in which publishing is highly dependent on supporting one's hypotheses, behaviors that lead to *appearing* correct are selected over behaviors that lead to actually *being* correct. Second are two related findings. In such a system, the Type I error rates and false discovery rates (i.e., the proportion of findings that are positive but incorrect) in the literature rise (see Figure 1). What this means is our literature, the "facts" we ideally teach our students, implement in our clinics, and use as the basis for further research, become increasingly likely to be incorrect. So, we are not only trying to find answers to some of the most complex questions of our time (Rahelić et al., 2024; Whitmarsh et al., 2021) with a game of 20 questions, but we are playing a rigged version. You see, in a game of 20 questions, it is assumed that

you can ultimately win because the answers you receive are correct; imagine playing a version of this game where this is not true, and it becomes increasingly not true as the game continues. This, it seems, is the game we all are playing today.

I think it is also important to pause here and consider two additional details addressed in Smaldino and McElreath (2016). First, the trends observed in the quality of the scientific literature were not caused by flaws in the character of the hypothetical scientists. Bad science was not the result of bad people but a flawed system—a concept we, as behaviorists, should understand well. Second, the safeguards we have put in place are not likely to have a significant impact if we choose to continue within this same system. Smaldino and McElreath (2016) found that even a very strong focus on replication, including increasing the number of replication studies over time and strong negative repercussions for one's results failing to replicate, had minimal effects. So, setting aside the ethics of inferring malicious intent where it need not exist and punishing individuals for a flawed system while continuing to support that system, even with the protections afforded by recent efforts focused on preregistration and replication, we likely will not make a significant impact on the type of work that gets produced. But this is not the only problem we face if we choose to continue down this road, for this system does not just reward being right. As mentioned previously, findings in this system are only important when one can frame one's work as innovative and novel, and this brings its own set of problems.

The Toothbrush Problem

Psychologists treat other people's theories like toothbrushes—no self-respecting person wants to use anyone else's. (Mischel, 2008)

Exactly as predicted by learning theory, a system that rewards novelty begets novelty (Smaldino & McElreath, 2016), or at least claims of novelty and innovation. I would argue, however, that *true innovation* is risky in such a system, as entering uncharted territory comes with fewer guarantees of being correct. How then does one produce innovative work without the risk inherent in doing such work? It would seem our system allows us to have our cake and eat it too because of a more insidious issue: our lack of focus on theory. By focusing on effects rather than the underlying explanations of what is driving those effects, we can continuously lay claim to novelty and

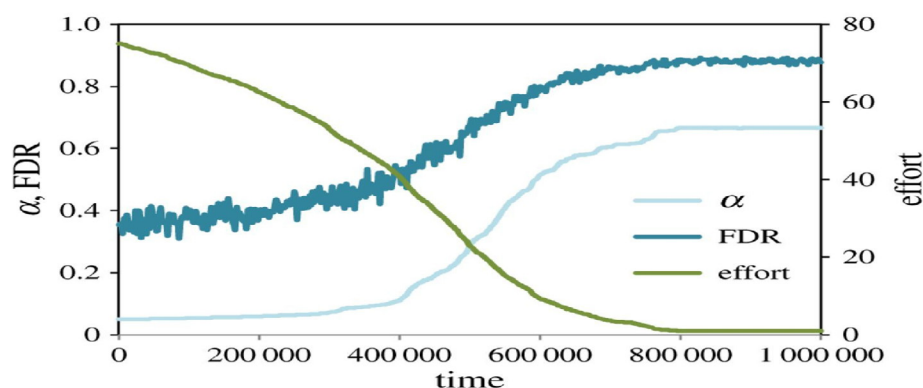


FIGURE 1 The evolution of science given current contingencies. Note. Reprinted from “The Natural Selection of Bad Science,” by P. E. Smaldino and R. McElreath, 2016, *Royal Society Open Science*, 3(9), p. 11 (<https://doi.org/10.1098/rsos.160384>). Copyright 2016 by the Authors.

innovation while doing nothing truly novel or innovative. We can change superficial features, come up with new names for our constructs and treatments, and in this way, we create the illusion, for ourselves as well as others, of novelty. In other words, we can each have our own toothbrush in the form of “new” treatments that are hailed as innovative and the next revolution in CBT when it is not entirely clear that anything other than the superficial features or the packaging of treatment components has changed. Perhaps this is why, over the course of my career, although the number of evidence-based treatments has grown significantly (see <https://www.psychologicaltreatments.org>), our data still seem to suggest that even our best treatments continue to help about 40–60% of people (Springer et al., 2018). So, although this system seems to work well for us as we build our careers, laying out a clear path to success, does it serve the public good? Is it helping us to serve our primary mission to alleviate human suffering? Are we pushing our science ever forward or are we simply reproducing what we already have in updated packaging? Perhaps most importantly, what would the alternative, a system focused more on the how and why of what we do rather than just the what, look like; would it be any better?

Theory: What Is It and How Can It Help?

As discussed previously, theories are explanations of what happens in the real world. To achieve such a complex aim, theory building, rather than focusing on the production of novel effects, strives to continuously integrate and accumulate what we know, synthesizing this knowledge into an increasingly broader understanding. An important focus of theory is on explaining the *established* facts in our field. Facts that have been shown to exist

independent of the methods (i.e., auxiliary hypotheses) that we use to establish their existence.⁴

In stark contrast to NHST, in which the goal is to be correct, as models of reality, good theories will always be false and incomplete. Whereas failure to support one’s hypothesis in NHST means one’s work won’t see the light of day, in theory building, flaws and errors are important building blocks toward progress. Theories capitalize on the scientific adage that we seem to have forgotten with generations of NHST—we learn more from our failures than our successes (see Haeffel, 2022, for a broader discussion of this point).

Finally, good theory is both broad and precise. This means that good theory provides an explanation about how a phenomenon works, not how it works in 18- to 25-year-old, White males; but at the same time, this explanation should define the exact nature of relationships. Relating to this last point, ideally good theory is formalized in the precise language of mathematics. Given that this type of theory building is so different from how many of us have been trained to work, I turn now to an example.

BUILDING A PRECISE THEORY OF PANIC DISORDER

Many members of ABCT will be familiar with a common verbal theory for panic disorder on which many of our treatments rely: Clark’s theory of panic disorder (depicted in Figure 2). Consistent with my arguments here, Clark’s theory has served us well, enabling us to develop treatments guided

⁴ To be clear, my definition here is not meant to exclude moderated effects—effects whose existence or magnitude varies under some condition(s)—moderated effects can be shown to exist reliably, independent of the specific methods used.

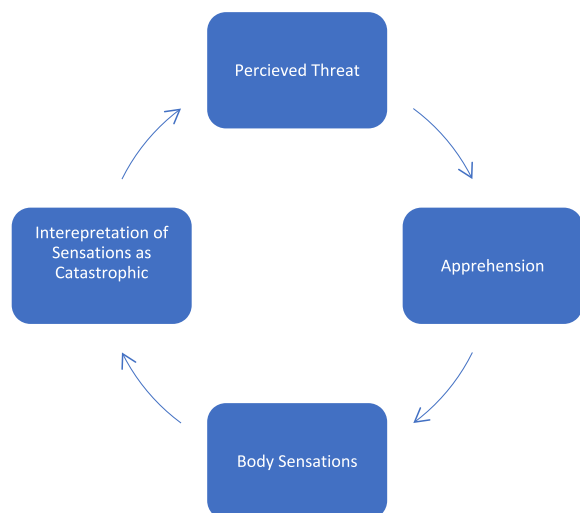


FIGURE 2 Clark's cognitive model of panic attacks. *Note.* Adapted from "A Cognitive Approach to Panic," by D. M. Clark, 1986, *Behaviour Research and Therapy*, 24(4), p. 463 ([https://doi.org/10.1016/0005-7967\(86\)90011-2](https://doi.org/10.1016/0005-7967(86)90011-2)). Copyright 1986 by Elsevier Ltd. Reused with permission.

by a shared understanding of the mechanisms that maintain panic disorder. However, verbal theories are limited by the imprecise nature of our language. Not infrequently over the course of my career, I have found myself in discussions with colleagues about the predictions made by a given verbal theory only to realize that each one of us was interpreting these theories somewhat differently. In fact, as we reacted to our lack of certainty by digging further into the literature to find the "right" interpretation, we found even in the published literature, verbal theories are interpreted differently depending on who is doing the interpretation or even when—the same researchers might operationalize a theory differently over time. A case in point, Robinaugh et al. (2021) offered four examples of how Clark's verbal theory could be formalized (see Figure 3).

Looking more closely at each of these possible interpretations, panel A of Figure 3 depicts a theory in which perceived threat has a direct and linear effect on arousal and vice versa. Alternatively, as depicted in panel B, it is possible that while perceived threat has a linear effect on arousal, low levels of arousal have a negligible effect on perceived threat until a critical point is reached, then within a particular range, increases in arousal result in rapid increases in perceived threat, with the change leveling off after another critical point is reached. Both of these possibilities are consistent with Clark's verbal theory of panic—either could be consistent with what Clark imagined when describing his theory; and to be clear, the four

examples offered by Robinaugh et al. (2021) in Figure 2 are only some of the possible ways that Clark's theory could be more precisely operationalized. One could easily imagine many other possibilities.

Looking at the first column in the rightmost panel in Figure 3 (theory-implied behavior), having specified more precise relationships between arousal and perceived threat, data simulations allow us to see what should happen under each theory if there is a discrete increase in arousal: for example, if one were to become out of breath after a jog. In the second column, we see what should happen when arousal varies naturally throughout the day. Under conditions described by theory A, a sudden increase in arousal would quickly result in a moderate increase in perceived threat, with the level of threat perception staying fairly steady during the remainder of the time period depicted (about 25 minutes). Theory A also suggests that natural variations in arousal experienced over the course of the day would result in variable levels of perceived threat throughout the day, with some repeated peaks and valleys, most remaining in a moderate range. Thus, based on what we know about the experience of panic attacks, we can conclude theory A is likely not to describe the nature of the relationship between threat and arousal that leads to the prototypical panic attack. The relationships described under theory B, however, suggest that a sudden increase in arousal would result in a quick, rapid rise in perceived threat, followed by a rapid decrease, with the rise and return to baseline taking place within a span of approximately 20 minutes. With normal variations in arousal over the course of the day, we see two sudden, sharp spikes in threat perception. Clearly, then, theory B better captures the real-world phenomenon of panic attacks. By more clearly specifying the relationships implicated in Clark's verbal theory of panic, and by using simulations to examine each, we can more precisely determine the likely interplay between threat perception and physiological arousal that gives rise to panic attacks and panic disorder.

Advancing our theories around psychopathology and its treatments has several benefits for those who suffer from these disorders. In terms of public health implications, one obvious benefit is that once a broad enough formal model is established, one could predict which events could be expected to lead to spikes in a disorder, guiding the deployment of needed resources. Just as importantly, more precise theories could allow us to reassure the families and policymakers when adverse events are not likely to lead to the stable

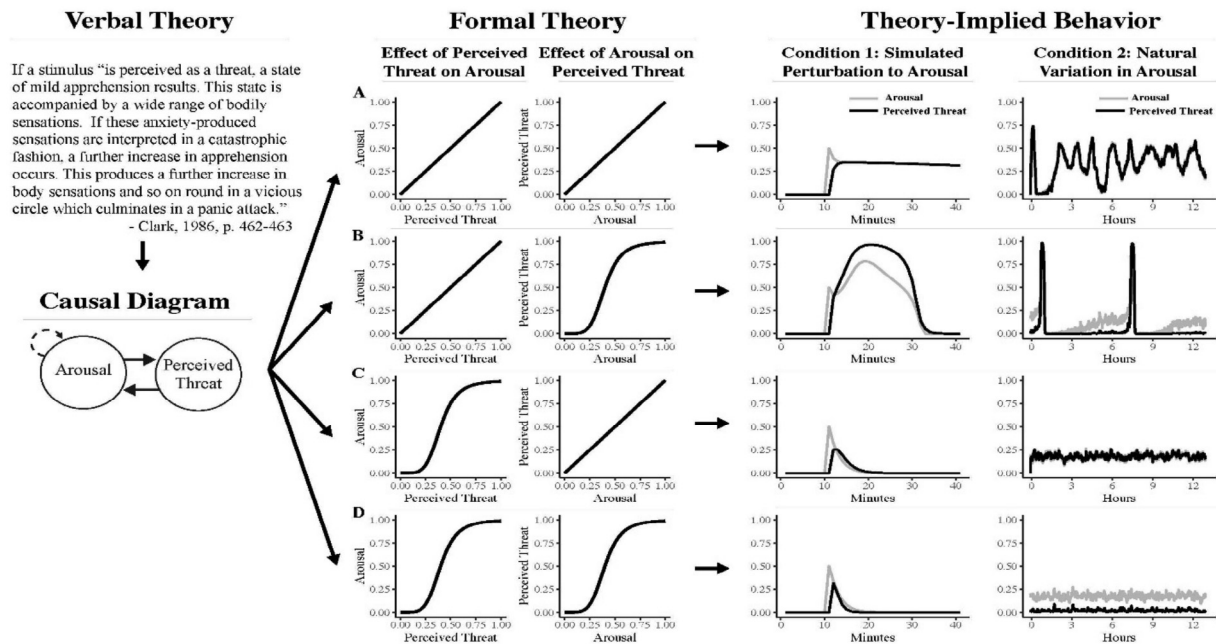


FIGURE 3 Formalization of Clark's verbal theory of panic. Note. Reprinted from "Invisible Hands and Fine Calipers: A Call to use Formal Theory as a Toolkit for Theory Construction," by D. J. Robinaugh, J. M. B. Haslbeck, O. Ryan, E. I. Fried and L. J. Waldorp, 2021, *Perspectives on Psychological Science*, 16(4), p. 729 (<https://dx.doi.org/10.1177/1745691620974697>). Copyright 2021 by the Authors. Reprinted with permission.

impairment and distress associated with psychological disorders. Additionally, one could easily see how advances in theory could lead to different treatments or to refinements in treatments that more precisely target underlying mechanisms. For example, support for the theory that normal variations in arousal that occur during the day increase threat perception, and ultimately panic attacks, might lead us to focus our attention on treatments that augment exposure therapy with interventions to decrease daily arousal levels (e.g., biofeedback, relaxation training, medication, or lifestyle changes). Alternatively, the support for the (still very basic) theory depicted in panel B can be used to guide clinical decision-making (e.g., implementing exposure practices that focus on violating outcome expectancies rather than habituation or decreases in arousal) and innovations in panic disorder treatment would be those that intervene either to decrease spikes in arousal or new methods that effectively break the link between sharp increases in arousal and threat perception.

Importantly, strong theoretical work in CBT has the potential to help us better understand the full range of topics relevant to improving our treatments. For example, staying with the given example, a broad and precise theory of the etiology and maintenance of panic disorder would like also lead to a better understanding of what therapist factors are related to prognosis in treatment, how to

define panic disorder, how or whether to differentiate it from disorders with substantial overlap in symptoms, and how to more accurately assess panic disorder and monitor treatment outcome. Moreover, relevant theories do not need to be tied to specific disorders but instead could focus on a myriad of topics relevant to advancing our science (e.g., a theory explaining variability in therapist outcomes in evidence-based treatments).

My point here is not to suggest that verbal theories are inherently bad or without merit. In fact, I believe, given the dearth of theory-guided research and practice in our field, any theoretical work should be applauded. However, verbal theories have several weaknesses that deserve significant attention. The first, I believe, is obvious from the previous discussion. Given the inherent imprecision of language, verbal theories will often be vague. This has several implications, with the most important being that vague theories are generally impervious to falsification. This is not just an abstract concern. Falsifiability is central to the scientific method because the knowledge gained from ruling out a possibility far exceeds the knowledge gained from ruling in a possibility. Additionally, the imprecision in verbal theories can make them impractical when attempting to make predictions under new conditions (e.g., mental health reactions to a pandemic). This can be true even of theories that have gained considerable empirical

support. The third issue is perhaps a less obvious problem, but one that I think is central to ABCT's toothbrush problem. The lack of precision in verbal theories can be understood as lack of transparency between the theorist and the rest of the scientific community. Often to understand what a verbal theory suggests under certain circumstances, one must either guess or have access to the theorist. Only the theorist themselves could state the theory implied behavior with absolute authority. In this way, verbal theories are inherently proprietary.

These are precisely the conditions that lead to Mischel's lamentation about psychology's troubled history with theory building—the original toothbrush problem. By developing a theory, particularly one impervious to contradictory findings, one can build a name for oneself in the field. This may be good for us personally, but can the same be said for the benefits to science and the public whom our science should serve? What about when we also consider that the vagueness of our theories allows the next person in line to come up with their own variation—whether this “advance” represents nonmeaningful differences in the terms ascribed to the constructs under consideration or more substantive changes is often not clear? Basing our treatments on vague theories allows us the same benefits (or leaves us open to the same pitfalls, depending on your perspective) as basing our treatment on no theory at all—what I refer to as ABCT's toothbrush problem. Defining a treatment as based on the specific activities that the therapist and client engage in during treatment, without a commensurate level of attention to innovation in how or why the treatment produces its effects, allows us to produce new treatment packages and programs, building our brand and adding to the choices available under the umbrella of evidence-based treatments, without knowing if we have made any substantive advances. However, if our innovations are merely “cosmetic,” different terms for the same construct, targeting the same mechanism of change in the same way and to the same degree with a superficially different intervention, we could expect that we would not see significant growth in the number of people who benefit from our treatments—this situation is not unlike what has been observed in our literature for several decades.

A Science Built on Precise Theory: Implications for the Public Good

How then, could more formal theory address these issues? Key differences between formalized theory and NHST are summarized in [Table 1](#), and below,

I discuss the implications of each of these differences as they related to producing science with the potential to translate into meaningful returns for the public.

Transparency and precision, a core feature of formalized theory (see [Table 1](#)), allow us as a field to move the focus to the science itself, as opposed to the scientist. With transparent ways of expressing our theory, theories are no longer propriety, under the authority of a single scientist or small group. The exact nature of the relationships among constructs of theory are apparent to the entire community of scientists and theory implied behavior under never-before-considered conditions can be modeled by anyone with the necessary skills. Importantly, this allows the scientific community to work collaboratively toward advancing the theory. Just as important, success in this environment is not about deriving one's one own propriety theory; instead, the focus shifts to making meaningful contributions to existing theory, the type of contributions that allow for increasingly broader explanations. Further, this shift in focus away from the scientist and toward our science allows us to become more objective and to embrace the advantages of identifying errors and limitations of any given theory. Focusing on where a theory fails, where we are wrong, is consistent with the scientific method and necessary to meaningful advances, but this perspective is at odds with how we critique, or often fail to critique, our work today.

The shift away from novelty also redefines success as the integration of knowledge into increasingly broader and more useful explanations. When this happens against the backdrop of precisely defined theories about how our treatments work, it is less likely that we can make claims of innovation without true advancement. Thus, a system that relies on formal theory is more likely to reinforce actual innovation as opposed to the appearance of innovation, bringing our contingencies in line with those that benefit the public good.

Finally, transparent and precise formal theories allow us to model what happens under new or emerging conditions. We saw in very clear terms what this meant in terms of the development of a vaccine for the novel coronavirus and in the predictions of how the virus would spread and the resulting morbidity and mortality. We also saw, however, the errors that can result from our reliance on less well-integrated and precise theory. As [Haefel \(2022\)](#) recounts in his own call for greater theory in psychological science, in 2021 the *American Psychologist* published a call to action in response to the mental health challenges

Table 1

Key Differences Between the Culture That Develops Around Null Hypothesis Significance Testing, Verbal Theory, and Formalized Theory

	NHST Culture	Verbal Theory	Mathematical/Computational Theories
What type of findings are valued?	Rejecting the null	Supporting the theory	Both theory consistent and theory inconsistent findings are useful
What types of topics are valued?	Novel	Accumulation and Integration of Knowledge	Accumulation and Integration of Knowledge
What level of precision is required?	High precision	Typically low	High
What is the level of transparency and ownership?	Low/Lab or Investigator	Low/Theory developer	High/Scientific Community
Generalizability	Low	Unclear	High
Applicability to Novel situations	Low	Low	High

Note. NHST = null hypothesis significance testing.

predicted to result from the COVID-19 pandemic (Gruber et al., 2021). This article, cited over 400 times to date, puts forth a credible argument that COVID-19, as a widespread, unpredictable, multi-dimensional stressor that requires behaviors (e.g., social distancing) limiting one's ability to use effective stress coping strategies, would result in both an acute and long-term need for increased mental health services, delivered in novel ways, and at a scale that could not be matched by our current workforce of mental health specialists alone. While this all seems to be highly plausible, and consistent with our current verbal theories of stress and coping, the fact is that these predictions were largely incorrect (see Haefel, 2022 Table 1, p. 5).⁵ In other words, we had years of data on stress and coping, we also had theory to guide our understanding, but the informal and imprecise nature of our theory requires us to make predictions based largely on clinical judgment, with all of the limitations and ethical implications that come with these types of predictions (Dawes, 2005; Dawes et al., 1989).

What Can We Do?

Given the centrality of theory building to the scientific endeavor and the potential benefits of formal theory, one might question why it is so rarely a focus in our field. I think there are some straightforward reasons as well as some that are more complex. First, many of us received little to no training in how to develop, test, or interpret formal theory. Our lack of familiarity with formal

theory building makes it hard to integrate into our own teaching, clinical work, or research. Moreover, another potential downside to formal theory building is it often will require collaboration between individuals with very different types of expertise—content experts (e.g., on the phenomenology and etiology of panic disorder) and experts in computational and mathematical modeling. Building such collaborations can be both time-consuming and challenging. This work is further complicated by the fact that, as models of reality, theory building requires the confluence of scientific expertise, clinical knowledge of how or whether laboratory findings are reliably manifested in the real world, and a sophisticated critical and philosophical analysis of what parts of reality are essential to model. In other words, there is no precise methodological roadmap that one can lay out that guarantees good theory and no shortcuts to in-depth and time-consuming analysis. Finally, formulation and dissemination of a precise formalized theory requires a comfort with being wrong—and being wrong in a transparent and public way.

In summary, theory building requires effortful, time-consuming behavior (i.e., lowered productivity in terms of the number of products), integration of established and reliable findings as opposed to focusing on novelty and innovation, being wrong, and transparency in being wrong. It is easy to see how this type of behavior is not elicited by the system in which we work today, one that, as I have argued here, selects for low-effort behaviors that result in decreased rigor but increased productivity, and positive findings (i.e., being right). So what can we do?

There are obvious solutions to many of the obstacles to theory building that I describe. It is (relatively) easy to build tools to connect

⁵ These comments are not meant to bely the assertion that there is a mental health crisis as one could argue that the levels of unaddressed or underaddressed mental health needs was at pandemic levels prior to the arrival of COVID-19.

collaborators. With the growing popularity of online learning and our ability to connect virtually, it is possible for graduate programs whose faculty do not have the requisite expertise in computational modeling to access experts throughout the world to offer courses or workshops for students. However, are these straightforward solutions enough to see real growth in a behavior that is not reinforced by the system in which we work? As a behaviorist, I find it hard to argue that these person-centered answers represent meaningful interventions if they are implemented without any concomitant change in the environment. Even if we ask our students, or task ourselves, with forgoing the reinforcers that come with success, [Smaldino and McElreath \(2016\)](#) have provided compelling evidence that this very act almost ensures one's inability to impact the field in the long term.

To make meaningful change, we must pivot away from the temptation to focus on the easy straightforward solutions or from telling ourselves it makes more sense to go for the low-hanging fruit "for now." I believe that if we are to develop a cumulative science that truly has the potential to bring about positive changes in our communities, we must fundamentally change what it means to be successful in our field today. We must ask ourselves the difficult questions. Does it make sense to continue the way we have? Have we driven real innovation by focusing on quantity—designating those among us who have the highest number of publications, or the most clinical hours when applying for internship as our success stories? Or, as I suggest here, do we need to focus on new metrics, metrics that examine whether our work as educators, clinicians, or researchers in cognitive behavioral science is making meaningfully impact in the real world?

What Can ABCT Do as an Organization?

ABCT's influence means, as an organization, we are in a unique position to support changes toward theory-driven CBT by mitigating both practical and cultural barriers. For example, when planning workshops, webinars, and the Annual Convention, ABCT could, and should, go outside the organization to bring members access to those who work specifically in theory development, computational modeling, and the philosophy of science. Similarly, the organization could support programs in providing needed training by putting out a call specifically for teaching resources on these topics to be disseminated through the ABCT website. To

directly stimulate more immediate advances, the organization could sponsor regular hack-a-thons that bring together content area experts and those with modeling expertise. Equally important, ABCT publications could introduce our members to the clinical relevance of theory-building efforts. Special issues of the organization's journals would serve to advance the dissemination of theoretical work as well as introduce the topic to many members who are unfamiliar with this type of work.

ABCT also has an important role to play in taking concrete steps to redefine success at both the individual and organizational level. As just one example, a report by the interorganizational task force *initiated by ABCT* to develop guidelines in CBT doctoral education emphasized that "Training in science and ethics form the backbone of education in CBP [cognitive and behavioral psychology]" ([Klepac et al., 2012, p. 691](#)), going so far as to include a list of topics in philosophy of science to which programs should introduce their students. Many of these topics would be key for developing students' understanding of how theory should influence their work in research, teaching, and clinical practice. In the context of growing demands on programs, however, it is unclear the degree to which this advice is being heeded. An early step for ABCT should be to review programs based on the [Klepac et al. \(2012\)](#) recommendations and to maintain a list of programs that are in substantial compliance with these recommendations. Moreover, the task force recommendations should certainly be central to the determination of which programs are honored by ABCT's awards program. Similarly, ABCT should examine the criteria across their awards program to ensure that both the criteria that selections committees use and the materials submitted by applicants. Align with values that promote meaningful advances in science over the quantity of one's output. Finally, although there is growing acknowledgment of the adverse consequences that have followed from the overreliance on metrics such as the h-index and journal impact factors (see for example, [Carleton et al., 2012](#)), there has been little effort to create environmental contingencies that correct the current trend. Given our roots in the science of understanding how contingencies shape behavior, ABCT should lead the way toward change by commissioning a task force that conquers the difficult problem of creating metrics or evaluation methods that are as easy to use as an h-index but that favor high quality work that produces meaningful advances in our science over high quantity output.

Conclusions

These are difficult goals and challenging problems to solve, but there is no better place to be having this conversation than at ABCT. Our membership is made up of approximately 5,000 cognitive behavioral scientists.⁶ While these numbers on their own are impressive given our humble beginnings—10 to 12 people meeting in a New York City apartment (Schere, 2016)—they bely the true impact of our members. ABCT members are some of the most passionate and driven individuals working in mental health and behavior change today. We count among our members those who drive the science and practice of CBT and those who determine what today's students, and thereby tomorrow's researchers, practitioners, and educators, learn about CBT. Our members direct the clinics where the public is most likely to receive CBT, they lead the educational programs in which students and trainees learn what CBT is today, and they conduct the research that will shape what CBT looks like tomorrow. Our senior members define what success is within the field by choosing which students get admitted to graduate programs and internships, who gets hired, and whose research gets funded and published. In terms we are all familiar with, our members are the drivers of the environmental contingencies, the rewards and punishments, that determine the trajectory of our field. Therefore, it is worth our time to pause to reflect on what exactly we have been prioritizing as a field and to question whether continuing down this path is right or is it simply comfortable. Failure to do so, I believe, represents an existential threat to the future relevance of CBT and, perhaps more importantly, it represents a present-day threat to those whom our work is supposed to serve.

References

- Bickman, L. (1996). A continuum of care: More is not always better. *American Psychologist*, 51(7), 689–701. <https://doi.org/10.1037/0003-066X.51.7.689>.
- Carleton, R. N., Parkerson, H. A., & Horswill, S. C. (2012). Assessing the publication productivity of clinical psychology professors in Canadian Psychological Association-accredited Canadian psychology departments. *Canadian Psychology/Psychologie Canadienne*, 53(3), 226–237. <https://doi.org/10.1037/a0027731>.
- Centers for Disease Control (2021). *WISQARS Leading Causes of Death Visualization Tool*. [https://wisqars.cdc.gov/lcd/?o=LCD&y1=2021&y2=2021&ct=12&cc=](https://wisqars.cdc.gov/lcd/?o=LCD&y1=2021&y2=2021&ct=12&cc=ALL&g=00&cs=0&r=0&ry=0&ce=0&ar=lcd1age&at=groups&ag=lcd1age&a1=0&a2=199)
- Cooper, M., & Kelland, H. (2015). Medication and psychotherapy in eating disorders: Is there a gap between research and practice? *Journal of Eating Disorders* 3, 45. <https://doi.org/10.1186/s40337-015-0080-0>.
- Cox, K. S., & Codd, R. T. III. (2024). Advocates of research-supported treatments for PTSD are losing in lots of ways: What are we going to do about It? *Research on Social Work Practice*, 34(4), 347–359. <https://doi.org/10.1177/10497315231206754>.
- Dawes, R. M. (2005). The ethical implications of Paul Meehl's Work on comparing clinical versus actuarial prediction methods. *Journal of Clinical Psychology*, 61(10), 1245–1255. <https://doi.org/10.1002/jclp.20180>.
- Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, 243(4899), 1668–1674. <https://doi.org/10.1126/science.2648573>.
- Deloitte Access Economics (2020). *The Social and Economic Cost of Eating Disorders in the United States of America: A Report for the Strategic Training Initiative for the Prevention of Eating Disorders and the Academy for Eating Disorders*. <https://www.hsph.harvard.edu/striped/report-economic-costs-of-eating-disorders/>.
- Eysenck, H. J. (1959). Learning theory and behaviour therapy. *Journal of Mental Science*, 105, 61–75. <https://doi.org/10.1192/bjp.105.438.61>.
- Forsyth, J. P., & Hawkins, R. P. (1997). Introduction to the special issue thirty years of Behavior Therapy: Promises kept, promises unfulfilled. *Behavior Therapy*, 28(3), 327–331. [https://doi.org/10.1016/S0005-7894\(97\)80077-9](https://doi.org/10.1016/S0005-7894(97)80077-9).
- Fried, E. I. (2020a). Lack of theory building and testing impedes progress in the factor and network literature. *Psychological Inquiry*, 31(4), 271–288. <https://doi.org/10.1080/1047840x.2020.1853461>.
- Fried, E. I. (2020b). 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>.
- Goldfried, M. R., & Davison, G. C. (1994). *Clinical behavior therapy* (expanded edition). John Wiley & Sons.
- Gruber, J., Prinstein, M. J., Clark, L. A., Rottenberg, J., Abramowitz, J. S., Albano, A. M., Aldao, A., Borelli, J. L., Chung, T., Davila, J., Forbes, E. E., Gee, D. G., Hall, G. C. N., Hallion, L. S., Hinshaw, S. P., Hofmann, S. G., Hollon, S. D., Joormann, J., Kazdin, A. E., ... Weinstock, L. M. (2021). Mental health and clinical psychological science in the time of COVID-19: Challenges, opportunities, and a call to action. *American Psychologist*, 76(3), 409–426. <https://doi.org/10.1037/amp0000707>.
- Haefel, G. J. (2022). Psychology needs to get tired of winning. *Royal Society Open Science*, 9(6), 220099. <https://doi.org/10.1098/rsos.220099>.
- Haefel, G. J., Thiessen, E. D., Campbell, M. W., Kaschak, M. P., & McNeil, N. M. (2009). Theory, not cultural context, will advance American psychology. *American Psychologist*, 64(6), 570–571. <https://doi.org/10.1037/a0016191>.
- Hofmann, S. G., Kasch, C., & Reis, A. (2025). Effect sizes of randomized-controlled studies of cognitive behavioral therapy for anxiety disorders over the past 30 years. *Clinical Psychology Review*, 117, 102553. <https://doi.org/10.1016/j.cpr.2025.102553>.
- Hofmann, S. G., Sawyer, A. T., & Fang, A. (2010). The empirical status of the “new wave” of cognitive behavioral therapy. *Psychiatric Clinics of North America*, 33(3), 701–710. <https://doi.org/10.1016/j.psc.2010.04.006>.

⁶ At times, and in my opinion, regrettably, the terms scientist and researcher have been used interchangeably; that is not my intent here. I use the term scientist to refer to those who approach cognitive behavioral therapy from a scientific perspective.

- Kazdin, A. E. (1978). *History of behavior modification: Experimental foundations of contemporary research*. University Park Press.
- Klepac, R. K., Ronan, G. F., Andrasik, F., Arnold, K. D., Belar, C. D., Berry, S. L., Christoff, K. A., Craighead, L. W., Dougher, M. J., Dowd, E. T., Herbert, J. D., McFarr, L. M., Rizvi, S. L., Sauer, E. M., & Strauman, T. J. (2012). Guidelines for cognitive behavioral training within doctoral psychology programs in the United States: Report of the Inter-organizational Task Force on Cognitive and Behavioral Psychology Doctoral Education. *Behavior Therapy*, 43(4), 687–697. <https://doi.org/10.1016/j.beth.2012.05.002>.
- Lash, T. L. (2017). The harm done to reproducibility by the culture of null hypothesis significance testing. *American Journal of Epidemiology*, 186(6), 627–635. <https://doi.org/10.1093/aje/kwx261>.
- 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>.
- Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science*, 34(2), 103–115. <http://www.jstor.org/stable/186099>.
- Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry*, 1(2), 108–141. https://doi.org/10.1207/s15327965pli0102_1.
- Mischel, W. (2008). The toothbrush problem. *Observer*, 21(11).
- Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual information processing*. Academic Press.
- Rahelić, V., Perković, T., Romić, L., Perković, P., Klobučar, S., Pavić, E., & Rahelić, D. (2024). The role of behavioral factors on chronic diseases-practice and knowledge gaps. *Healthcare*, 12(24). <https://doi.org/10.3390/healthcare12242520>.
- Robinaugh, D. J., Haslbeck, J. M. B., Ryan, O., Fried, E. I., & Waldorp, L. J. (2021). Invisible hands and fine calipers: A call to use formal theory as a toolkit for theory construction. *Perspectives on Psychological Science*, 16(4), 725–743. <https://doi.org/10.1177/1745691620974697>.
- Schere, M. L. (2016). 50 years ago: Who's been hanging out in Dorothy's apartment? The humble origin of the Association for Advancement of Behavior Therapy. *The Behavior Therapist*, 39(7), 222–228.
- Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science*, 3(9), 160384. <https://doi.org/10.1098/rsos.160384>.
- Spencer, M. R., Garnett, M. F., & Miniño, A. M. (2024). *Drug overdose deaths in the United States, 2002–2022*. National Center for Health Statistics.
- Springer, K. S., Levy, H. C., & Tolin, D. F. (2018). Remission in CBT for adult anxiety disorders: A meta-analysis. *Clinical Psychology Review*, 61, 1–8. <https://doi.org/10.1016/j.cpr.2018.03.002>.
- Weisz, J. R., Han, S. S., & Valeri, S. M. (1997). More of what? Issues raised by the Fort Bragg study. *American Psychologist*, 52(5), 541–545. <https://doi.org/10.1037/0003-066X.52.5.541>.
- Whitmarsh, L., Poortinga, W., & Capstick, S. (2021). Behaviour change to address climate change. *Current Opinion in Psychology*, 42, 76–81. <https://doi.org/10.1016/j.copsyc.2021.04.002>.
- Wolitzky-Taylor, K., Zimmermann, M., Arch, J. J., De Guzman, E., & Lagomasino, I. (2015). Has evidence-based psychosocial treatment for anxiety disorders permeated usual care in community mental health settings? *Behaviour Research and Therapy*, 72, 9–17. <https://doi.org/10.1016/j.brat.2015.06.010>.

RECEIVED: April 24, 2025

REVISED: August 19, 2025

ACCEPTED: August 29, 2025

AVAILABLE ONLINE: 3 SEPTEMBER 2025