



Published in final edited form as:

*Personal Sci.* 2021 ; 2: . doi:10.5964/ps.6001.

## Why Has Personality Psychology Played an Outsized Role in the Credibility Revolution?

Olivia E. Atherton<sup>1</sup>, Joanne M. Chung<sup>2</sup>, Kelci Harris<sup>3</sup>, Julia M. Rohrer<sup>4</sup>, David M. Condon<sup>5</sup>, Felix Cheung<sup>6</sup>, Simine Vazire<sup>7</sup>, Richard E. Lucas<sup>8</sup>, M. Brent Donnellan<sup>8</sup>, Daniel K. Mroczek<sup>1,9</sup>, Christopher J. Soto<sup>10</sup>, Stephen Antonoplis<sup>11</sup>, Rodica Ioana Damian<sup>12</sup>, David C. Funder<sup>13</sup>, Sanjay Srivastava<sup>5</sup>, R. Chris Fraley<sup>14</sup>, Hayley Jach<sup>7</sup>, Brent W. Roberts<sup>14,15</sup>, Luke D. Smillie<sup>7</sup>, Jessie Sun<sup>16</sup>, Jennifer L. Tackett<sup>9</sup>, Sara J. Weston<sup>5</sup>, K. Paige Harden<sup>17,18</sup>, Katherine S. Corker<sup>19</sup>

<sup>[1]</sup>Department of Medical Social Sciences, Northwestern University, Chicago, IL, USA.

<sup>[2]</sup>Department of Psychology, University of Toronto, Mississauga, Canada.

<sup>[3]</sup>Department of Psychology, University of Victoria, Victoria, Canada.

<sup>[4]</sup>Department of Psychology, Leipzig University, Leipzig, Germany.

<sup>[5]</sup>Department of Psychology, University of Oregon, Eugene, OR, USA.

<sup>[6]</sup>Department of Psychology, University of Toronto, Toronto, Canada.

<sup>[7]</sup>School of Psychological Sciences, University of Melbourne, Melbourne, Australia.

<sup>[8]</sup>Department of Psychology, Michigan State University, East Lansing, MI, USA.

<sup>[9]</sup>Department of Psychology, Northwestern University, Evanston, IL, USA.

<sup>[10]</sup>Department of Psychology, Colby College, Waterville, ME, USA.

<sup>[11]</sup>Department of Psychology, University of California, Berkeley, CA, USA.

<sup>[12]</sup>Department of Psychology, University of Houston, Houston, TX, USA.

<sup>[13]</sup>Department of Psychology, University of California, Riverside, CA, USA.

This is an open access article distributed under the terms of the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/), CC BY 4.0, which permits unrestricted use, distribution, and reproduction, provided the original work is properly cited.

**Corresponding Author:** Olivia E. Atherton, Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, 625 N. Michigan Avenue, 22nd floor, Chicago, IL, USA 60611. olivia.atherton@northwestern.edu.

**Author Contributions:** *Olivia E. Atherton*—Writing | Feedback, revisions. *Joanne M. Chung*—Writing | Feedback, revisions. *Kelci Harris*—Writing | Feedback, revisions. *Julia M. Rohrer*—Writing | Feedback, revisions. *David M. Condon*—Writing | Feedback, revisions. *Felix Cheung*—Writing | Feedback, revisions. *Simine Vazire*—Writing | Feedback, revisions. *Richard E. Lucas*—Writing | Feedback, revisions. *M. Brent Donnellan*—Writing | Feedback, revisions. *Daniel K. Mroczek*—Writing | Feedback, revisions. *Christopher J. Soto*—Feedback, revisions. *Stephen Antonoplis*—Feedback, revisions. *Rodica Ioana Damian*—Feedback, revisions. *David C. Funder*—Feedback, revisions. *Sanjay Srivastava*—Feedback, revisions. *R. Chris Fraley*—Feedback, revisions. *Hayley Jach*—Feedback, revisions. *Brent W. Roberts*—Feedback, revisions. *Luke D. Smillie*—Feedback, revisions. *Jessie Sun*—Feedback, revisions. *Jennifer L. Tackett*—Feedback, revisions. *Sara J. Weston*—Feedback, revisions. *K. Paige Harden*—Feedback, revisions. *Katherine S. Corker*—Writing | Feedback, revisions.

**Supplementary Materials:** Materials [see Index of Supplementary Materials]

**Competing Interests:** Brent Donnellan, David Funder, Brent Roberts, Julia Rohrer, Luke Smillie, and Simine Vazire are members of the editorial board of the journal.

**Ethics Statement:** No ethical issues and/or ethics approvals need to be disclosed.

**Related Versions:** No other previously published versions of this manuscript exist in part or in whole.

[14]Department of Psychology, University of Illinois, Champaign, IL, USA.

[15]Hector Research Institute of Education and Sciences and Psychology, University of Tübingen, Tübingen, Germany.

[16]Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA.

[17]Department of Psychology, University of Texas, Austin, TX, USA.

[18]Population Research Center, University of Texas, Austin, TX, USA.

[19]Department of Psychology, Grand Valley State University, Allendale, MI, USA.

## Abstract

Personality is not the most popular subfield of psychology. But, in one way or another, personality psychologists have played an outsized role in the ongoing “credibility revolution” in psychology. Not only have individual personality psychologists taken on visible roles in the movement, but our field’s practices and norms have now become models for other fields to emulate (or, for those who share Baumeister’s (2016, <https://doi.org/10.1016/j.jesp.2016.02.003>) skeptical view of the consequences of increasing rigor, a model for what to avoid). In this article we discuss some unique features of our field that may have placed us in an ideal position to be leaders in this movement. We do so from a subjective perspective, describing our impressions and opinions about possible explanations for personality psychology’s disproportionate role in the credibility revolution. We also discuss some ways in which personality psychology remains less-than-optimal, and how we can address these flaws.

## Keywords

personality psychology; credibility revolution; meta-science; replication crisis

---

Social psychology might think carefully about how much to follow in personality psychology’s footsteps. Our entire field might end up being one of the losers.

— Baumeister (2016, p. 158)

When it comes to its role in the credibility revolution, personality psychology is punching above its weight. To an interested observer watching the unfolding of this movement, it might be difficult to make sense of personality psychology’s outsized role. Some of the most visible advocates for reform have come from personality psychology, despite the fact that we are a very small, and typically not-very-influential, subfield. Moreover, the stereotype of personality research is that it is boring, not especially successful at attracting attention, funding, or respect, and does not share some of the core features of its larger neighbors (e.g., it is rarely experimental like social and cognitive psychology typically are). Some may find it curious that such a minor subfield of psychology has played a disproportionate role in this movement. We do, too. We can understand the feeling, which seems to underlie Baumeister’s warnings in the paper quoted above, that personality psychology is not the subfield you’d expect other subfields to look to for leadership.

The authors of this paper share a curiosity about why personality psychology has had such an outsized role in the credibility revolution. We do not know the answer. However, we believe that sharing our experiences, speculations, and perceptions of our field may help others better understand this period in psychology's history. In the short-term, we hope to stimulate discussions and alternative explanations for this curious phenomenon. In the long-term, we hope our reflections may provide grist for scholars (e.g., historians, philosophers, and sociologists) studying psychology's replication crisis and credibility revolution.

This paper should not be considered a definitive or authoritative investigation of the causes of personality psychology's outsized role. Instead, it should be seen as a subjective account from people on the front lines, reporting their experiences and venturing some guesses about possible explanations. In many places, we chose to present opinions and observations that are not backed up by empirical data in the interest of providing a richer discussion. Any claim presented without a citation or supporting evidence should be treated as speculative and not as fact.

As with most multi-authored opinion pieces, not all authors share all perceptions and opinions articulated here. What we share is an interest in the question of *why* personality psychology may have played a disproportionate role in the credibility revolution and some experiences that speak to possible answers. Our experiences are quite diverse—we have different (sometimes conflicting) intellectual commitments; we are at different stages of our careers and so have witnessed different parts of personality psychology's history and from different vantage points; we have demographic differences that shape our experiences; and we have different biases, values, and personalities.

We organize this paper into two broad sections: strengths and weaknesses. First, we elaborate on some of the unique strengths of personality psychology and possible explanations as to why we may have taken on the credibility revolution in stride. Specifically, we draw on our history and norms as a field, as well as the methodological and statistical practices that have long been common in personality psychology. Then, in the second section of this paper, we turn to discussing issues that undermine our credibility as a field -- notably, methodological and statistical practices, causal inference, and diversity and generalizability. While our role in the credibility revolution suggests we can set a positive example in some domains, we should not ignore the fact that we have some glaring negative qualities as a field. Thus, this paper captures both optimistic and critical tones about the state of personality psychology's credibility and our outsized role in the revolution. What factors put personality psychology in a position to play such a role in the credibility revolution? What are some of the pitfalls that personality psychology has not overcome? Where do we most need to improve?

## **Strengths of Personality Psychology: What We Do Well**

### **History and Field Norms**

Some of the factors that may have led us to be well-prepared for the credibility revolution stem from events in our history and the cultural norms we have as a field. One episode in our history that left a noticeable mark and arguably set us up for success in the face of

the replication crisis, is another crisis that forced us to justify our credibility as a scientific (versus pseudoscientific) field: the person-situation debate. Of course, there are many non-parallels between the person-situation debate and the replication crisis (including the validity of the criticisms leveled against the field), but there are also many striking parallels. For this reason, we suspect that the lessons learned from the person-situation debate played an important role in setting us up to deal with the replication crisis and become involved in the credibility revolution.

The person-situation debate questioned the existence of personality traits, and whether personality versus situational factors mattered more in determining behavior (Lucas & Donnellan, 2009; Mischel, 1968). Thus, the very existence of personality psychology as a field was at stake, at least in the United States. Importantly, personality's crisis in the 1970s and 1980s in the US was different from the current replication crisis (and many other so-called "crises" that other subfields have faced) in one important way: it threatened the existence of the subfield. The possibility that graduate programs and faculty jobs would disappear, journals would fold, and funding would dry up was very real (e.g., Baumeister, 1999; Swann & Seyle, 2005) - indeed, these developments were under way before things started to turn around in the mid 1980s. Nevertheless, the person-situation debate has many similar features to psychology's current replication crisis. Most pertinently, many psychologists are now questioning whether everything we've been studying is just noise or whether we are studying something real.

For both the person-situation debate and the credibility revolution, these existential questions drew widespread attention and subsequent arguments over how seriously these concerns should be taken (Epstein & O'Brien, 1985; John et al., 2012). Furthermore, discussions about both the person-situation debate and methodological reform were (and are) rather heated. Raw feelings over the person-situation debate have lasted decades (e.g., Lucas & Donnellan, 2009), and critics of the credibility revolution have gone so far as to make public derogatory remarks about scientists interested in methodological reform (e.g., "shameless little bullies," "replication police," "methodological terrorists").

In both cases, the debates over the credibility of the field forced researchers to justify the field's credibility on scientific grounds. This forced difficult (and rare) conversations about the quality and integrity of research practices and findings, and prompted researchers to formulate competing (and testable) hypotheses regarding the central issues at hand. For example, the person-situation debate led personality scientists to craft ways to empirically test hypotheses about whether personality traits are real (e.g., via cross-situational consistency) and the relative impact of personality traits (vs. situations) on behavior (e.g., Kenrick & Funder, 1988). Likewise for the credibility revolution, critics voiced their dissent using arguments that were subsequently turned into testable hypotheses such as the impact of researcher expertise on the success of replication attempts (Many Labs 4; Klein et al., 2019) and the role of subjective analytic choices on research findings (Silberzahn et al., 2018).

Personality psychologists may have been primed to play an outsized role in the credibility revolution in part because of this sense of familiarity with the revolution we've already

been through, and because many of the norms and values promoted via the credibility revolution are practices that we have long adopted (likely, at least in part, because of the person-situation debate). The credibility revolution aims to promote a culture where individual scientists and findings are not put on pedestals; where healthy, open critical discourse and incremental research are necessary for scientific progress; where data sharing and replications are valued; and where conclusions about the data are tempered accordingly. Thanks to our experience having to justify our credibility in the wake of the person-situation debate, personality psychologists have engaged with many of these practices before they were cool.

For better and for worse, personality psychologists have mostly failed to gain public attention (and also status/attention within psychology), despite our subject matter being of great interest to the general public, and undeniably important to social and political events. Our inability to translate the popularity of our subject matter into status or influence may be due in part to our struggles establishing our credibility as a field and our place in psychology. This lack of status meant we had fewer opportunities to put individual scientists on pedestals, which has proven to be helpful in the age of the credibility revolution because criticisms and evaluations of personality research were less likely to be seen as personal attacks. Moreover, the fact that personality psychology is a relatively small field, has little money at stake (in terms of grants or external partnerships), isn't popular within scientific psychology, and largely lacks access to status and attention may have made it so that we generally had less to lose when the replication crisis hit.

Because status is not a major barrier to criticism in our field, this may have also led us to adopt a general culture of strong disagreements within the field, aired quite openly but without a lot of hostility or personal animosity (at least in the last few decades). For example, personality scientists subscribe to different views regarding personality structure and measures (e.g., Big Five, HEXACO, individual trait approaches such as the Q-sort; Anglim & O'Connor, 2019; Wiernik et al., 2020) and have differing opinions about what the most useful levels and units of analysis are (e.g., nomothetic vs. idiographic). Despite these disagreements, personality psychologists cooperate in their scientific pursuits, and critical discourse is carried out through empirical research.

This culture of disagreements and open scientific exchange has arguably led to important advances. Personality psychologists seem to value incremental research, place high value on measurement validity, and have less of a fixation on novelty, thus making us more "boring" (Baumeister, 2016; Funder, 2016). Furthermore, personality psychologists were collecting larger samples, sharing data, reporting null effects, and replicating across studies before these became widely-accepted as being best practices across the rest of psychology. Personality psychologists have also historically valued calibrating conclusions to the data, which is possibly due to intense scrutiny by critics during the person-situation debate.

Post person-situation debate, some may wonder whether personality psychology succeeded in establishing its credibility. From our perspective, it seems as though we succeeded well enough to survive but not well enough to become a popular subfield, or to be well-accepted into the broader field. Time will tell if the credibility revolution and the field of

metascience will face the same fate. It is noteworthy that, in the pursuit of establishing credibility, personality psychologists and meta-scientists both developed formal societies. These represent small (but growing) efforts to solidify our respective (and hopefully permanent) places in the field.

### Methods and Statistical Validity

Our long history of “boringness” with respect to methods and statistical validity has proven, in some ways, to pay off, setting us up to meet the replication crisis with confidence. Many of the popular avenues for methodological reform in the age of the credibility revolution are methodological practices that personality psychology has long been obsessed with: 1) psychometrics to establish the existence and validity of traits, 2) using large sample sizes, and 3) a dedication to descriptive and exploratory research.

The field had an early focus on psychometrics, in part because it was necessary. In order to measure traits and develop trait taxonomies, it was critical to take the steps to establish construct validity. Are we measuring what we intend to measure? Cattell (1946) was one of the first to employ factor analysis in the study of traits, a technique that proved formative in the derivation of the Big Five and HEXACO trait models. Additionally, concepts and tools such as convergent and discriminant validity (Loevinger, 1957), the nomological network (Cronbach & Meehl, 1955), and the multi-trait, multi-method matrix (Campbell & Fiske, 1959) have their origins in individual differences research, and these concepts continue to be not only in heavy use in personality today, but have proven to be invaluable tools for robust measurement and rigorous research across many subfields.

Concepts like the nomological network and the multi-trait, multi-method matrix also encourage personality psychologists’ use of multiple methods—including self-report, informant-report, behavioral observation, and biological measures—to triangulate on a claim. For instance, Harari and colleagues (2020) used mobile sensing techniques to examine how personality traits are associated with objective measures of behavior in everyday life. In the area of personality neuroscience, examples abound of researchers triangulating biological measures against self-report or behavioral measures (e.g., Jach et al., 2020). Although the field is sometimes rightly critiqued for an overreliance on self-report questionnaires, the frequent use of multiple methods is notable. Certainly, each method has strengths and weaknesses, but checking the consistency of results across methods have helped the field form a more robust and credible evidence base. These practices serve as a model for other fields seeking to increase their credibility.

Another strength of personality psychology is our track record of using observational data with relatively large sample sizes. A handful of studies have found that research areas and studies that use primarily observational, as opposed to experimental, methods tend to use larger sample sizes (Fraley & Vazire, 2014). In addition, Kossmeier et al. (2019) found a steady increase in sample size over time in two top personality journals. These larger samples may be partly due to more frequent use of statistical techniques that require large samples (e.g., exploratory factor analysis, structural equation modeling), or due to the sometimes lower costs of data collection (e.g., cross-sectional surveys), or due to personality researchers’ tendency to use longitudinal panel studies like the German Socio Economic



Panel (GSOEP) or other large existing datasets. Regardless of the reason, larger samples are important for the credibility of research findings because they increase statistical power, thus reducing the incidence of false negatives and increasing the incidence of true positives (which also reduces the false discovery rate, or the proportion of positive findings that are false positives). Additionally, higher powered studies produce more stable effect size estimates (Schönbrodt & Perugini, 2013), which reduces the chances of stumbling on fluke findings and exacerbating publication bias.

Finally, to show the real-world relevance of personality traits, researchers rely on ecologically-valid exploratory and descriptive studies to test the distributions, antecedents, and consequences of traits. This focus on ecologically-valid, descriptive research (which blends confirmatory and exploratory approaches) possibly reduces pressure for personality psychologists to *p*-hack. For instance, some researchers have conducted descriptive studies to document how personality traits vary across nations (e.g., Rentfrow et al., 2013) and across the life course (e.g., Lucas & Donnellan, 2011), and how they predict important life outcomes such as physical health, subjective well-being, and work performance (e.g., Ozer & Benet-Martinez, 2006). Plus, the adoption of broad personality taxonomies, such as the Big Five, makes the file drawer less of a problem and publishing null results much easier (i.e., even if only one of the Big Five traits correlates with the outcome, we can still publish the null effects of the other four traits). For example, researchers may test hypotheses derived from social investment theory regarding how conscientiousness changes when young adults enter the workforce. These confirmatory goals are complemented by an exploration of how the remaining Big Five traits, in addition to conscientiousness, change following joining the workforce. Null results are reasonably common when looking at the antecedents and consequences of five different traits.

Through all of these examples, we can see how our focus on a variety of methodological techniques serves to improve not only reproducibility but also credibility and robustness. A focus on measurement and construct validity provides a solid foundation from which to study personality. A focus on large samples improves statistical inferences. A focus on ecological validity and descriptive research enhances the generalizability and applicability of our research findings and reduces the file drawer problem. These features have led us to be well-prepared to meet the demands of the credibility revolution.

## **Weaknesses of Personality Psychology: There's Still Work to Do**

### **Methods and Statistical Validity**

The many relative strengths of personality as a discipline do not preclude the need for improvement. Although our strengths are heavily concentrated in the domain of methods and statistics, some of our biggest challenges are also in this area, specifically with respect to measurement, design, analytics, and effect size interpretation. For personality psychology to continue to lead in this area, a recognition of the ways that methodological issues affect the practice of personality science is critical.

One notable example is a heavy reliance on verbal reports and especially self-reports to measure core constructs. Although the use of multi-method data is becoming increasingly

common, there are a non-trivial number of personality studies that rely on self-reports for both independent and dependent variables. This monomethod design produces artificially inflated effect sizes. Even very small amounts of shared method variance can lead to “replicable confounding” that does not reflect substantively important associations. Moreover, personality psychology is plagued by jingle-jangle fallacies, where sometimes different constructs go by the same name (jingle) and other times, the same constructs go by different names (jangle). Confusion over construct terminology markedly hinders scientific progress in the field (e.g., Leising et al., 2020).

Another particularly salient challenge is related to study design and the reuse of existing data, which presents many opportunities but also has potential pitfalls. Personality psychologists often rely on secondary data analyses of existing, large, longitudinal, and nationally-representative datasets, which means that many key design choices are outside the researcher’s control, such as the selection of measures. Thus, the investigator’s judgment is required to create measures from existing item pools, and the scope of available choices provides ample opportunities for flexibility. When combined with the fact that using large-scale studies often requires choices about which subsamples to include, which waves to analyze, and which covariates to model, there can be a staggering number of decisions to make, thus providing a fertile ground for capitalizing on chance. Moreover, many of these datasets are reused for different purposes by the same or different research teams. Prior knowledge can inform subsequent analyses in ways that facilitate confirmation bias and capitalize on sample-specific characteristics (see Weston et al., 2019, for ways to address these challenges).

Another way the complexity of personality research threatens the accumulation of robust empirical findings is that some longitudinal research designs are difficult to reproduce because of the money and time involved, and thus are rarely replicated. The credibility revolution has shed light on the importance and challenges of replication (Zwaan et al., 2018), primarily because attempts to replicate seemingly simple laboratory experiments failed. Replication is even more complicated when applied to complex personality research. For example, it is exceedingly rare to have longitudinal studies with the exact same measures, data collection intervals, and sampling plans. The apparent lack of commensurability of measurement and design features makes the task of replicating key findings from complex studies all the more difficult. Issues with robustness could go unnoticed due to the practical reality that direct replications of some research designs in personality psychology are hard to conduct.

Moreover, it is not always clear what defines a successful replication. Although debates about how to evaluate whether a replication is successful impact all research areas (Zwaan et al., 2018), this challenge is even greater when studies do not focus on a single, dichotomous test of statistical significance or single effect size estimate. For instance, age-related trajectories in personality traits are thought to be replicable across studies; however, trajectories can be distinguished by general patterns or by their specific shape, age at peak, and more (Graham et al., 2020) and impressionistic evaluations of replicability may gloss over important inconsistencies. Similarly, many critical debates in the field center around support for specific factor analytic or structural equation models, and evaluation of these



models typically relies on subjective judgments about model fit and the general similarity of coefficients across studies. In short, many of the analyses that personality psychologists conduct necessarily require judgment and subjectivity to interpret, which means that replicability can be overstated. Collectively, these factors threaten the accumulation of robust empirical findings in personality research.

A related concern is that many complicated statistical models are based on assumptions and data-dependent subjective decisions that are not always transparent to readers and reviewers. And often, it is difficult to pre-register analytic plans involving complex modeling because it is nearly impossible to anticipate all of the problems that could arise while analyzing the data. The fact that researchers cannot easily anticipate all of the choices they will need to make and problems they will encounter introduces flexibility despite their best intentions. Researchers often need to constrain certain parameters to get complex models to converge, and whether such constraints are justifiable is open to debate. These problems are addressable with open data, well-documented code, and thorough transparent reporting. However, many of these practices are not yet widely adopted. It is important to distinguish statistical complexity from rigor.

Finally, beyond the double-edged sword of sophisticated modeling approaches, other methodological features in personality research can also have subtle downsides. For instance, the move away from null hypothesis significance testing toward estimating and evaluating effect sizes is a strength of personality psychology (e.g., Funder & Ozer, 2019). However, there is no consensus on what constitutes a meaningful effect size for personality research. This is due, in part, to the fact that the connection between basic research in personality and application to the real-world is often lacking, and also because personality researchers do not attend to effect sizes as much as they should. It is safe to say that most personality effects are not large. This reality is understandable given that many outcomes of practical importance are multiply determined, so researchers should not expect any single variable to be strongly associated with any outcome. However, this fact creates challenges in interpreting whether small and reliable effects of personality are theoretically and practically important, an issue that personality psychologists will need to grapple with. What distinguishes a small but important effect from one that is negligible?

### **Causality and Internal Validity**

Related to these issues of measurement, study design, complex analytic approaches, and effect sizes is another issue that personality psychologists historically (and presently) grapple with: causality. Personality psychologists often rely on observational designs, as many research questions do not readily lend themselves to randomized experiments. For example, personality psychologists are often interested in the causal effects of traits on behaviors and outcomes, but we lack the means to randomly assign individuals' personality.<sup>1</sup> Additionally, we are interested in determinants of personality (e.g., social class, birth order position, maturation) that cannot be directly manipulated either. This raises doubts about the internal validity of our findings.

---

<sup>1</sup>)Although no random assignment of personality is currently possible, other designs can approximate such manipulations. For example, individuals may be randomly assigned to express higher or lower levels of specific traits (Jacques-Hamilton et al., 2019).

Although many other subfields of psychology have a single-minded focus on experiments as the only acceptable means to establish causality (and maybe even as the only way to do “proper” research), this is also problematic because, depending on the object of inquiry, experiments may not be the best way to proceed (Rozin, 2001). Experiments do indeed guarantee internal validity under very weak assumptions, but they do not guarantee broader valid causal inferences because of concerns about external validity.

Instead of trying to emulate a strictly experimental mode of science, it may be more instructive for personality psychologists to pay close attention to other disciplines that rely on observational data. For example, epidemiologists tackle many research questions that cannot be solved through experimentation alone, such as whether or not smoking causes lung cancer (Pearl & Mackenzie, 2018), as do economists, political scientists, and sociologists. These fields seem to share a greater acceptance of the goal of causal inference on the basis of observational data, and a stronger reliance on formalized frameworks such as Directed Acyclic Graphs (e.g., Rohrer, 2018) or the Potential Outcomes Framework (Rosenbaum, 2017). Furthermore, they more frequently make use of difference-in-difference designs, instrumental variables, and other study types that can be classified as natural experiments, which further strengthens causal inferences (e.g., Damian et al., 2021; Dunning, 2012). Likewise, the field of behavioral genetics offers genetically-informative designs, which can strengthen causal inferences considerably (e.g., Briley et al., 2018). Personality psychologists are in an excellent position to learn from these adjacent fields and incorporate causal inference as a serious scientific endeavor rather than something that should be hidden behind weaselly language (Grosz et al., 2020).

Nevertheless, taking causal inference seriously will require us to overhaul current practices. For example, while confounding variables are almost universally identified as an important problem, little awareness seems to exist about how to properly select control variables, or how sampling affects causal inference (Rohrer, 2018). Personality psychologists’ enthusiasm for longitudinal data may prove to be a unique strength because such data can improve causal inferences if employed properly (see VanderWeele et al., 2016). However, some overhauling of statistical models may be required as well. While certain classes of models are mainly used for descriptive purposes (such as visualizing trajectories via growth curves), cross-lagged panel models (CLPM) and their modifications (such as the random intercept CLPM; Hamaker et al., 2015) have often been taken as tools to establish causality, despite the fact that they require additional strong assumptions (Usami et al., 2019).

There is no consensus on whether and how to draw causal inferences from observational data, and causal inference generally poses challenges that should not be understated. However, if personality psychologists are willing to accept this challenge and to catch up on developments from other disciplines, we may be able to contribute to yet another “credibility revolution” (such as the one in economics; Angrist & Pischke, 2010), introducing proper causal inference to observational research in psychology.

### **Diversity and Generalizability**

Even if we were to address the methodological and statistical issues discussed above, the field of personality psychology would continue to suffer from another serious threat to our

credibility: the homogeneity of perspectives represented in our field. Diverse perspectives are vital to good science for a number of reasons. First, diversity in who is represented contributes to diversity in the field's substantive foci; *who* is doing the research impacts *what* research is being conducted (Roberts et al., 2020). A more diverse group of scholars will likely study a broader range of research questions, use a more complete set of tools, and include a more diverse pool of participants. For a field that prizes methodological triangulation, as personality psychology does, a lack of diversity is a serious threat to the comprehensiveness of our research programs. Second, diversity is critical for a robust, self-critical, and self-correcting scientific community (Longino, 1990). Researchers bring their background experiences and assumptions to their own work and to their critical appraisal of others' work. If personality research is only read and critiqued by a homogenous group of researchers who share similar biases and blind spots, we are less likely to identify and correct flaws in our work. A serious commitment to self-correction, and therefore to credibility, requires a commitment to creating an environment where a diverse range of scholars are included and valued. Personality psychology is failing at this important task.

One major threat to diversity is that the field has often been unwelcoming to women and underrepresented minorities (URMs) - groups whose percentages in our personality undergraduate programs, graduate programs, faculty positions, and professional societies are lower than their percentages in the general population and whose voices have historically not been given the opportunity to be heard. It is our sense that personality psychology contains a highly disproportionate number of researchers who identify as men and White. Specific demographic data are not collected by some of our major societies, such as the Association for Research in Personality (ARP) and the European Association of Personality Psychology (EAPP), suggesting that the demographics of our members either have been considered irrelevant and unnecessary, or have not been considered at all. The lack of data also makes it difficult to track whether efforts made towards equity and inclusion are working. This situation is both a symptom of personality psychology's past and a major liability to its future. The most overt examples of hostility towards women and underrepresented minorities stem from norms of professional behavior that are common to psychology more broadly, which allow harassment and discriminatory behavior to perpetuate in many professional contexts, including at scientific meetings, in hiring and promotion practices, and with respect to issues affecting scientific prestige.

Furthermore, personality science has historically centered the experiences of White men, extending back to the origins of psychological assessment. Prior to World War II, the most widely-used personality assessments (e.g., Woodsworth's 1917 Psychoneurotic Inventory and House's 1927 Mental Hygiene Inventory) contributed to substantial selection biases in military and workplace screenings (Gibby & Zickar, 2008), all while early intelligence tests were weaponized in favor of blatantly racist, discriminatory policies (e.g., *Buck v. Bell*, 1927). Since then, the field has adopted more inclusive methods for assessing "normative" personality and a broader acknowledgement of the probabilistic relationship between traits and behavioral outcomes. Yet, group differences in psychological traits are still sometimes reported with inadequate or inappropriate interpretations (e.g., J. P. Rushton's work), raising the possibility for substantial harm by promulgating stereotypes, not only to members of

the described groups but also to members of the research community who find such claims offensive.

Relatedly, another problem for generalizable inferences is methodological complacency. For instance, the Big Five measurement model has, for good reasons, been readily embraced by the field, so much so that variants of the Big Five and Big Six frameworks are now used almost exclusively. This persists despite evidence that fails to support the universality of these dimensions (De Raad et al., 2010), much less their true cross-cultural relevance (Laajaj et al., 2019). Although consensus can be a strength by minimizing the degree to which competing camps dig their heels in, it may also be a cause of methodological homogeneity (cf. Thalmayer et al., 2020).

Another double-edged sword is the use of panel study data. These datasets allow researchers to save time and money, and give them access to large nationally representative samples, but personality psychologists have overwhelmingly focused on analyzing panel data from Western countries (e.g., the German Socio-Economic Panel [SOEP], Midlife in the US, and Dutch Longitudinal Internet studies for the Social Sciences [LISS] datasets), despite the existence of several non-Western panel datasets (e.g., Midlife in Japan Study, the Indonesian Family Life Study, the Chinese Family Panel Study, the Korean Longitudinal Survey of Women and Families, and the sister studies to the Health and Retirement Study collected across the globe).

Moreover, as of October 2020, a review of the 50 most cited empirical papers that list personality as a keyword indicates that all 50 papers were authored by people with institutional affiliations in the United States, Canada, Germany, the UK, and New Zealand, and only three papers included samples outside of these regions (see Supplementary Materials). Interestingly, only one paper title mentions the sociodemographic characteristics of the people being sampled, and the other titles generally refer to psychological constructs rather than people, consistent with the idea that, “One consequence of a USA-centric sampling bias in psychology may be biased assumptions of (White) people from the United States as especially reflective of humankind” (Cheon et al., 2020, p. 1).

A step in the right direction would be to acknowledge the self-perpetuating features of these concerns. An attainable first move would be for our professional organizations to document member demographics (e.g., gender, racial and other identities, nationality, etc.). This would provide a clearer sense of the diversity of people who are currently in the field, as well as a benchmark to assess our progress in the future. This is, however, a very basic first step. It is also important to build on this initial step by making this information accessible and transparent to members of the field, so that in our professional society and departmental contexts, we can make informed decisions about what steps need to be taken to diversify. In other words, collecting the data is not enough; it must be accessible, transparent, and turned into action to have an impact.

Another, perhaps more difficult, endeavor is to reflect on the culture that we may be perpetuating in our classes, laboratories, departments, and research output. For example, in our teaching and knowledge mobilization, we could take care to avoid characterizing certain

traits as wholly adaptive (or maladaptive) without communicating that social and contextual factors are necessary for understanding what is considered “adaptive” or not (e.g., delay of gratification is not “adaptive” and “desirable” in all social and cultural contexts; Watts et al., 2018). Likewise, we could emphasise the fact that personality traits are not “fixed”. People change in important ways, and these malleable individual differences are critical for understanding social issues, as has been shown in the behavioral genetic literature at the intersection of genetic risk, education, and social mobility (e.g., Herd et al., 2019).

Furthermore, we could consider the extent to which we can improve the spaces we inhabit as personality psychologists to make them more welcoming towards people who are underrepresented in the field. One necessary endeavor is to increase URM scholarship in personality psychology and to amplify the voices of URM groups that have historically not been given the opportunity to provide input. This could be encouraged by including URM students in our research activities at the undergraduate level by hiring them as paid research assistants or making use of work-study programs. We can also invite URM scholars to join our editorial boards, present at our conferences, and be our collaborators. In doing so, we can work toward lessening the divide between spaces that are occupied by people who are oriented towards what we might consider “mainstream” personality psychology, and topics that are often investigated by URM scholars (e.g., research on identity may be seen as more central to personality psychology than is ethnic identity research). It is also important that we not only make concerted efforts to invite URM scholars to spaces that majority-group scholars dominate, but to also connect with URM scholars in their spaces (e.g., by attending [currently virtual] conferences such as the American Arab Middle Eastern and North African Psychological Association’s) to better understand what URM scholars find important in relation to personality research and what they need to feel supported in academia. We can continue virtual programming for societies, like ARP and EAPP, to facilitate more international participation and reduce resource barriers that members of underrepresented groups face in attending in-person conventions. We can educate ourselves with the myriad of anti-racism resources available on the Internet and participate in trainings offered by our institutions and within the online community (i.e., Academics for Black Survival and Wellness<sup>2</sup>). Until we fix the field’s diversity problem, personality psychology cannot reach its full potential as a credible, robust science.

## Conclusion

This is our perspective on how personality psychology may have been well-suited to play an outsized role in the credibility revolution, as well as the ways in which our field is in need of improvement. We might be wrong; there are many parts of our perspective that have caveats, “with notable exceptions”, biases, and blind spots. After all, the authors of this piece often had differing perspectives, which we have attempted to incorporate into this piece as much as possible. And, we acknowledge that the factors that set us up for success with the credibility revolution may not be the same factors that will set other fields up for success. There are, of course, individual differences in what works and what doesn’t.

---

<sup>2</sup>)See [www.academics4blacklives.com](http://www.academics4blacklives.com)

Some may wonder: what is the cost of increasing rigor? It has been noted that,

“Unquestionably the study of personality is far more rigorous today than it was half a century ago. Yet back then personality psychology captured the interest of people in many different fields, influencing anthropology, literary criticism, philosophy, plus lots of other subfields in psychology. The gain in rigor was accomplished by a loss in interest value... its influence on thinkers in other fields is far less, and indeed it has failed to capture the imagination of the intellectual community”

(Baumeister, 2016).

We may indeed be more boring, but from our perspective, what we gain from rigor, in terms of scientific truth, far outweighs that cost. The larger movement towards increasing scientific credibility reaches well beyond social and personality psychology, and even beyond psychology more broadly, to fields like biomedicine, physics, and ecology. If psychologists eschew rigor and fail to embrace transparency, the entire enterprise of psychological research may decline in its influence altogether, even to the point of irrelevance. We could end up being interesting but ultimately untrustworthy, re-invigorating early conceptions of psychology as a pseudo-scientific, rather than scientific, endeavor. To address this, psychologists must challenge themselves and each other to do better, to capitalize on our respective strengths and to acknowledge where we need to make change. We have attempted to do that here for our subfield, and we believe that with the collective efforts of all subfields, we can build a brighter and more credible future.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments:

We would like to thank Stefan Schmukle and Felix Schönbrodt for their feedback on early versions of this manuscript.

## Funding:

The authors have no funding to report.

## References

- Anglim J, & O'Connor P (2019). Measurement and research using the Big Five, HEXACO, and narrow traits: A primer for researchers and practitioners. *Australian Journal of Psychology*, 71(1), 16–25. 10.1111/ajpy.12202
- Angrist JD, & Pischke JS (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *The Journal of Economic Perspectives*, 24(2), 3–30. 10.1257/jep.24.2.3
- Baumeister RF (1999). On the interface between personality and social psychology. In Pervin LA & John OP (Eds.), *Handbook of personality: Theory and research* (pp. 367–377). Guilford Press.
- Baumeister RF (2016). Charting the future of social psychology on stormy seas: Winners, losers, and recommendations. *Journal of Experimental Social Psychology*, 66, 153–158. 10.1016/j.jesp.2016.02.003
- Briley DA, Livengood J, & Derringer J (2018). Behaviour genetic frameworks of causal reasoning for personality psychology. *European Journal of Personality*, 32(3), 202–220. 10.1002/per.2153



- Buck v. Bell, 274 U.S. 200 (1927). JUSTIA. <https://supreme.justia.com/cases/federal/us/274/200/>
- Campbell DT, & Fiske DW (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81–105. 10.1037/h0046016 [PubMed: 13634291]
- Cattell RB (1946). Personality structure and measurement. I. The operational determination of trait unities. *British Journal of Psychology*, 36(2), 88–103. 10.1111/j.2044-8295.1946.tb01110.x
- Cheon BK, Melani I, & Hong Y (2020). How USA-centric is psychology? An archival study of implicit assumptions of generalizability of findings to human nature based on origins of study samples. *Social Psychological & Personality Science*, 11(7), 928–937. 10.1177/1948550620927269
- Cronbach LJ, & Meehl PE (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302. 10.1037/h0040957 [PubMed: 13245896]
- Damian RI, Serrano S, & Hill PL (2021). Hurricane exposure and personality development. *Journal of Personality*, 89(1), 35–49. 10.1111/jopy.12542 [PubMed: 32031677]
- De Raad B, Barelds DPH, Levert E, Ostendorf F, Mla i B, Blas LD, H ebí ková M, Szirmák Z, Szarota P, Perugini M, Church AT, & Katigbak MS (2010). Only three factors of personality description are fully replicable across languages: A comparison of 14 trait taxonomies. *Journal of Personality and Social Psychology*, 98(1), 160–173. 10.1037/a0017184 [PubMed: 20053040]
- Dunning T (2012). *Natural experiments in the social sciences: A design-based approach*. Cambridge University Press.
- Epstein S, & O'Brien EJ (1985). The person-situation debate in historical and current perspective. *Psychological Bulletin*, 98(3), 513–537. 10.1037/0033-2909.98.3.513 [PubMed: 4080897]
- Fraley RC, & Vazire S (2014). The N-pact factor: Evaluating the quality of empirical journals with respect to sample size and statistical power. *PLOS ONE*, 9(10), Article e109019. 10.1371/journal.pone.0109019
- Funder D (2016, May 12). Why doesn't personality psychology have a replication crisis? [Blog post]. *funderstorms*. <https://funderstorms.wordpress.com/2016/05/12/why-doesnt-personality-psychology-have-a-replication-crisis/>
- Funder DC, & Ozer DJ (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, 2(2), 156–168. 10.1177/2515245919847202
- Gibby RE, & Zickar MJ (2008). A history of the early days of personality testing in American industry: An obsession with adjustment. *History of Psychology*, 11(3), 164–184. 10.1037/a0013041 [PubMed: 19048975]
- Graham EK, Weston SJ, Gerstorff D, Yoneda TB, Booth T, Beam CR, Petkus AJ, Drewelies J, Hall AN, Bastarache ED, Estabrook R, Katz MJ, Turiano NA, Lindenberger U, Smith J, Wagner GG, Pedersen NL, Allemand M, Spiro A... Mroczek D (2020). Trajectories of Big Five personality traits: A coordinated analysis of 16 longitudinal samples. *European Journal of Personality*, 34(3), 301–321. 10.1002/per.2259 [PubMed: 33564207]
- Grosz MP, Rohrer JM, & Thoemmes F (2020). The taboo against explicit causal inference in nonexperimental psychology. *Perspectives on Psychological Science*, 15(5), 1243–1255. 10.1177/1745691620921521 [PubMed: 32727292]
- Hamaker EL, Kuiper RM, & Grasman RP (2015). A critique of the cross-lagged panel model. *Psychological Methods*, 20(1), 102–116. 10.1037/a0038889 [PubMed: 25822208]
- Harari GM, Müller SR, Stachl C, Wang R, Wang W, Bühner M, Rentfrow PJ, Campbell AT, & Gosling SD (2020). Sensing sociability: Individual differences in young adults' conversation, calling, texting, and app use behaviors in daily life. *Journal of Personality and Social Psychology*, 119(1), 204–228. 10.1037/pspp0000245 [PubMed: 31107054]
- Herd P, Freese J, Sicinski K, Domingue BW, Mullan Harris K, Wei C, & Hauser RM (2019). Genes, gender inequality, and educational attainment. *American Sociological Review*, 84(6), 1069–1098. 10.1177/0003122419886550
- Jach HK, Feuerriegel D, & Smillie LD (2020). Decoding personality trait measures from resting EEG: An exploratory report. *Cortex*, 130, 158–171. 10.1016/j.cortex.2020.05.013 [PubMed: 32653745]
- Jacques-Hamilton R, Sun J, & Smillie LD (2019). Costs and benefits of acting extraverted: A randomized controlled trial. *Journal of Experimental Psychology. General*, 148(9), 1538–1556. 10.1037/xge0000516 [PubMed: 30489119]

- John LK, Loewenstein G, & Prelec D (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23(5), 524–532. 10.1177/0956797611430953 [PubMed: 22508865]
- Kenrick DT, & Funder DC (1988). Profiting from controversy: Lessons from the person-situation debate. *The American Psychologist*, 43(1), 23–34. 10.1037/0003-066X.43.1.23 [PubMed: 3279875]
- Klein RA, Cook CL, Ebersole CR, Vitiello C, Nosek BA, Ahn P, Brady AJ, Christopher R, Chartier CR, Christopherson CD, Clay S, Collisson B, Crawford JT, Cromar R, Gardiner G, Gosnell CL, Grahe J, Hall C, Howard I, Joy-Gaba JA... Ratliff KA (2019). Many labs 4: Failure to replicate mortality salience effect with and without original author *involvement*. PsyArXiv. <https://psyarxiv.com/vef2c/>
- Kossmeier M, Vilsmeier J, Dittrich R, Fritz T, Kolmanz C, Plessen CY, Slowik A, Tran US, & Voracek M (2019). Long-term trends (1980–2017) in the N-pact factor of journals in personality psychology and individual differences research. *Zeitschrift für Psychologie mit Zeitschrift für Angewandte Psychologie*, 227, 293–302. 10.1027/2151-2604/a000384
- Laajaj R, Macours K, Hernandez DAP, Arias O, Gosling SD, Potter J, Rubio-Codina M, & Vakis R (2019). Challenges to capture the Big Five personality traits in non-WEIRD populations. *Science Advances*, 5(7), Article eaaw5226. 10.1126/sciadv.aaw5226
- Leising D, Thielmann I, Glöckner A, Gärtner A, & Schönbrodt F (2020). Ten steps toward a better personality science: How quality may be rewarded more in research evaluation. PsyArXiv. <https://psyarxiv.com/6btc3/>
- Loevinger J (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, 3(3), 635–694. 10.2466/pr0.1957.3.3.635
- Longino HE (1990). *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton University Press.
- Lucas RE, & Donnellan MB (2009). If the person-situation debate is really over, why does it still generate so much negative affect? *Journal of Research in Personality*, 43(2), 146–149. 10.1016/j.jrp.2009.02.009
- Lucas RE, & Donnellan MB (2011). Personality development across the life span: Longitudinal analyses with a national sample from Germany. *Journal of Personality and Social Psychology*, 101(4), 847–861. 10.1037/a0024298 [PubMed: 21707197]
- Mischel W (1968). *Personality and assessment*. Psychology Press.
- Ozer DJ, & Benet-Martinez V (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology*, 57, 401–421. 10.1146/annurev.psych.57.102904.190127
- Pearl J, & Mackenzie D (2018). *The book of why: the new science of cause and effect*. Basic Books.
- Rentfrow PJ, Gosling SD, Jokela M, Stillwell DJ, Kosinski M, & Potter J (2013). Divided we stand: Three psychological regions of the United States and their political, economic, social, and health correlates. *Journal of Personality and Social Psychology*, 105(6), 996–1012. 10.1037/a0034434 [PubMed: 24128185]
- Roberts SO, Bareket-Shavit C, Dollins FA, Goldie PD, & Mortenson E (2020). Racial inequality in psychological research: Trends of the past and recommendations for the future. *Perspectives on Psychological Science*, 15(6), 1295–1309. 10.1177/1745691620927709 [PubMed: 32578504]
- Rohrer JM (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, 1(1), 27–42. 10.1177/2515245917745629
- Rosenbaum PR (2017). *Observation and experiment*. Harvard University Press.
- Rozin P (2001). Social psychology and science: Some lessons from Solomon Asch. *Personality and Social Psychology Review*, 5(1), 2–14. 10.1207/S15327957PSPR0501\_1
- Schönbrodt FD, & Perugini M (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, 47(5), 609–612. 10.1016/j.jrp.2013.05.009
- Silberzahn R, Uhlmann EL, Martin DP, Anselmi P, Aust F, Awtrey E, Bahník Š, Bai F, Bannard C, Bonnier E, Carlsson R, Cheung F, Christensen G, Clay R, Craig MA, Dalla Rosa A, Dam L, Evans MH, Flores Cervantes IF... Carlsson R (2018). Many analysts, one data set: Making

- transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*, 1(3), 337–356. 10.1177/2515245917747646
- Swann WB Jr., & Seyle C (2005). Personality psychology's comeback and its emerging symbiosis with social psychology. *Personality and Social Psychology Bulletin*, 31(2), 155–165. 10.1177/0146167204271591 [PubMed: 15619589]
- Thalmayer AG, Job S, Shino E, Robinson SL, & Saucier G (2020). † sigu: A mixed-methods lexical study of character description in Khoekhoegowab. *Journal of Personality and Social Psychology*; Advance online publication. 10.1037/pspp0000372
- Usami S, Murayama K, & Hamaker EL (2019). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods*, 24(5), 637–657. 10.1037/met0000210 [PubMed: 30998041]
- VanderWeele TJ, Jackson JW, & Li S (2016). Causal inference and longitudinal data: a case study of religion and mental health. *Social Psychiatry and Psychiatric Epidemiology*, 51(11), 1457–1466. 10.1007/s00127-016-1281-9 [PubMed: 27631394]
- Watts TW, Duncan GJ, & Quan H (2018). Revisiting the marshmallow test: A conceptual replication investigating links between early delay of gratification and later outcomes. *Psychological Science*, 29(7), 1159–1177. 10.1177/0956797618761661 [PubMed: 29799765]
- Weston SJ, Ritchie SJ, Rohrer JM, & Przybylski AK (2019). Recommendations for increasing the transparency of analysis of preexisting data sets. *Advances in Methods and Practices in Psychological Science*, 2(3), 214–227. 10.1177/2515245919848684 [PubMed: 32190814]
- Wiernik BM, Yarkoni T, Giordano CA, & Raghavan M (2020). Two, five, six, eight (thousand): Time to end the dimension reduction debate! *PsyArXiv*. <https://psyarxiv.com/d7jye/>
- Zwaan RA, Etz A, Lucas RE, & Donnellan MB (2018). Making replication mainstream. *Behavioral and Brain Sciences*, 41, Article e120. 10.1017/S0140525X17001972

**Relevance Statement**

In the future, philosophers of science, historians, and researchers alike may want to understand what it was like “living” through the credibility revolution in psychology and beyond. We provide our subjective view as personality psychologists.

**Key Insights**

- An account of personality psychology’s outsized role in the credibility revolution.
- Strengths include historical events, field norms, and methodological/statistical approaches.
- Weaknesses include methodological/statistical approaches, establishing causality, and diversity/generalizability.
- This is our subjective perspective.
- We might be wrong.