

Current Challenges and Proposed Solutions in Psychological Sciences



Onurcan Yılmaz¹  

¹ Kadir Has University, Faculty of Economics, Administrative and Social Science, Psychology, Istanbul, Türkiye

Abstract

The behavioral sciences, and particularly psychology, are currently grappling with a replication crisis. Moral psychology, like other fields, has not been spared from this challenge. This crisis can be attributed to several factors, including the inherent difficulties of research within these disciplines and the prevalent “publish or perish” culture in academia. However, some of these causes are within researchers’ control. Primary issues include the lack of overarching theoretical models to explain findings, insufficient standardization in measurement tools and statistical analysis methods, an overreliance on intention measures, and problems with representativeness. These limitations hinder the proper integration of cultural-evolutionary explanations and fail to account for complex interactions that underlie primary effects in the behavioral sciences. In this paper, I first provide a general overview of these challenges and then propose practical solutions for each. These suggestions include adopting the abductive method, prioritizing validity studies, implementing open science practices, recognizing the limitations of our inferences, enhancing representativeness, developing behavioral measures, gathering evidence from various disciplines, and promoting collaboration in large research teams.

Keywords

Reproducibility crisis • theory crisis • validity crisis • WEIRD • crisis • open science



“ Citation: Yılmaz, O. (2025). Current challenges and proposed solutions in psychological sciences. in S. Cesur, M. Karasu, B. Aktaş E., S. Balım, C. Çelikadam & B. Aktar (Eds.), *Social Psychological Insights into Current Morality Topics* (p. 22-33). İstanbul University Press. <https://doi.org/10.26650/B/SSc18.2025.015.002>

© This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International License. 

© 2025. Yılmaz, O.

✉ Corresponding author: Onurcan Yılmaz onurcan.yilmaz@khas.edu.tr

Behavioral sciences and psychology are currently grappling with a replication crisis. Although numerous factors contributing to this crisis have been identified, few studies have systematically and comprehensively examined these causes. It appears that the replication crisis arises from various interconnected sub-problems, and addressing each separately is essential for developing a comprehensive solution. In this section, firstly, I will outline the characteristics common to studies that have been successfully replicated. Secondly, I will explore several specific sub-problems contributing to the replication issue, including the theory crisis, the validity crisis, the inference crisis, the WEIRD (Western, Educated, Industrialized, Rich, and Democratic) crisis, the intention-behavior gap, and the subject matter of psychology. Following this analysis, I will provide practical suggestions for students and researchers on addressing these challenges, based on my own experiences.

The Replication Crisis: Which Studies Replicate Better?

Direct Replication

Replication studies are commonly divided into two types: direct and conceptual. In direct replication, the manipulation or data collection tools that operationally define the variables are replicated as closely as possible to the original study. In contrast, conceptual replication employs alternative measurement or manipulation tools to define the variable operationally. Cognitive psychologists tend to prioritize the robustness of their research through direct replication, while social psychologists tend to emphasize conceptual replication to ensure the generalizability of effects. Although specific data are limited, this distinction may explain why findings in cognitive psychology are generally more replicable than those in social psychology. Given the typically small effect sizes in psychology, it appears reasonable to first determine the robustness of an effect through direct replication. Therefore, if a study has been directly and successfully replicated, it is reasonable to assume that it will be easier to replicate in future experiments.

Theoretical Maturity

The theoretical maturity of the effect under replication should also be considered as well as direct replication. Theoretical maturity refers to a theory's ability to outline potential cause-effect relationships between variables and their boundary conditions prior to any data collection. While achieving this at the outset may not always be feasible, robust phenomena and their boundary conditions can often be empirically identified through direct or closely related replication studies that examine different boundary conditions of the issue. These findings can thereafter be integrated back into the theory to enhance its maturity.

Theories that fail to articulate a network of possible relationships between variables typically demonstrate low maturity. Experiments testing hypotheses derived from such underdeveloped theories are inherently more prone to replication failures due to the numerous potential boundary conditions under which the effects might occur. While it is empirically challenging to eliminate all alternative conditions, a mature theory can effectively minimize them. For example, in moral psychology, the Morality as Cooperation Theory (Curry et al., 2019) provides a comprehensive evolutionary perspective by linking morality to evolutionary processes and aligning moral values with cooperation. However, it does not adequately specify the conditions under which the theory could be falsified, particularly the proximate mechanisms involved. The lack of theoretical maturity to present such alternatives, coupled with a lack of methodological unity to test these possibilities consistently, hampers theoretical progress in these and similar fields (Meehl, 1967). Thus, theoretical maturity in moral psychology involves explaining not only the evolutionary history of morality across species but also detailing the psychological and



developmental mechanisms that influence the emergence of morality. Currently, moral psychology does not possess a unified theory that comprehensively addresses all these aspects.

Validity and Statistical Power

Attempting to replicate an experiment and failing to find a statistically significant effect does not necessarily mean the effect is unreliable or invalid (Maxwell et al., 2015). There are at least four possible reasons for a replication failure. Firstly, the manipulation used might be ineffective. If the effect size is small and the manipulation weak, the failure to detect the effect could simply be due to inadequate manipulation. Secondly, while the effect might be significant, a dependent variable consisting of scale items that remain unaffected by experimental manipulations (e.g., personality traits) may lack sufficient sensitivity to capture the effect. Thirdly, even with a strong manipulation and a sensitive dependent variable, insufficient statistical power can result in a failure to detect small effects. For example, to detect a small effect size in an independent variable with three levels, a sample of 1,548 participants is required to achieve 95% statistical power. If the sample size is considerably smaller, the primary reason for not finding an effect could simply be a lack of statistical power. Only after ruling out these three factors can we consider the fourth possibility: the effect is not robust and fails to be reliably replicated. In other words, robust effects are more likely to be replicable when the statistical power is high, the manipulation is strong, and the measurement of the dependent variable is highly sensitive.

Context of the Original Study and the Replication Study

Another crucial factor is the extent to which a replication study successfully mirrors the context of the original study. Psychological effects are not isolated from their context. Accordingly, replication studies that incorporate features of the original setting can more accurately identify the boundary conditions of the effect and verify the authenticity of the original finding. For instance, in a study we conducted (Isler et al., 2023), we attempted to replicate a well-known psychological effect (Shah et al., 2015) during the COVID-19 pandemic. The effect in question involved context-dependent pricing, in which, for example, an individual might be willing to pay \$20 for a beer at a concert they enjoy, whereas the same beer might only cost \$5 outside the concert setting. The original finding by Shah et al. (2015) suggested that as people's socioeconomic status decreases, they are less influenced by this context effect and would likely pay the same amount for the beer in both settings. During our replication, we measured perceptions of COVID-19 risk and found that the original effect did not emerge under the same experimental conditions; however, it did when accounting for COVID-19 risk perception. Thus, we were able to replicate the original study by considering a significant contextual difference between our replication setting and the original study's context. Had we not accounted for this, the replication might have failed, not necessarily indicating that the original effect is unreliable; rather, it may suggest that differences in design and context were not sufficiently considered or were overlooked.

A long-term strategy to mitigate replication failures involves maturing theories and deriving inferences directly from them; a short-term approach emphasizes using valid data collection tools and identifying and addressing potential contextual differences between the original study and its replication early on. Such differences may not always be apparent to the replicator, hence, a proactive approach involves contacting the original study's corresponding author to discuss the replication design and seek feedback. This approach proved beneficial in our research when an original author suggested measuring and controlling for COVID-19, which allowed us to identify a crucial boundary condition.

Open Data and Science

Replicability is likely enhanced in studies adhering to open data and open science practices, as minor differences can significantly impact replication outcomes. The original study should make all procedural details and specific measurement methods available. Until the 2020s, such transparency was not the norm in behavioral sciences, with researchers seldom sharing their data or full procedures openly. The open science movement has significantly contributed to improving this transparency (Nosek et al., 2022). For instance, researchers attempting to systematically replicate 193 cancer studies found that they could not access complete experimental protocols without additional inquiries, a challenge commonly observed in psychological studies as well (Errington et al., 2021).

Competence of the Researcher

Although there is no conclusive evidence suggesting that the competence of researchers impacts replication success (cf., Bench et al., 2017), which should not be overlooked. A lack of evidence does not equate to evidence of absence, particularly in failed replications. As a researcher in a non-Western culture, I have observed many instances where Turkish stimuli were poorly translated in cross-cultural studies, yet these studies were published in top journals. This issue often arises when responsibility is diffused among large groups of researchers or when there is insufficient incentive to meticulously work on multi-author papers. This underscores that the skill and diligence of the researcher may be critical in influencing the replication process.

Current cross-cultural replication efforts often involve open calls for participation, where participants are typically accepted as researchers. These participants collect data and subsequently engage in translation. However, this inclusive approach means that researchers might range from master's or PhD students to even undergraduates, potentially leading to variability in sample quality and unplanned errors. Researchers familiar with such studies recognize that this process is far from flawless. Validity studies, essential for ensuring the reliability of findings, are often overlooked in favor of direct hypothesis testing. Typically, existing scales developed in Western contexts are merely translated into other languages and used without thorough validation.

Considering these challenges, a logical assessment might suggest the importance of considering the scientific competence of replication groups and establishing criteria for participation in such research. By elevating the role of replication to a career step, individuals who choose to specialize in multi-laboratory (cross-cultural) studies might approach these projects with greater diligence. Additionally, it may become necessary for participants to demonstrate their capability to conduct high-quality research independently, as evidenced by relevant publications.

Just as method development is recognized as a career with prestigious journals dedicated to it, such as *Behavior Research Methods*, conducting high-quality replication studies and their analyses could also become a viable career goal. Developing incentive systems is crucial; by initially incentivizing researchers, followed by funders, and finally the journal system, significant changes can be gradually implemented. This approach mirrors the successful tactics of the open science revolution. Starting with incentives for pre-registration, the movement gained momentum as funders began requiring open science compliance, and journals increasingly favored practices like pre-registration and registered reports. By employing a similar strategy, we can explore alternative concepts such as careers in replication or theoretical psychology and devise strategies to establish norms that enhance the overall quality of scientific research.



Registered Report

The registered report is an innovative paper format born from the open science movement. In this format, researchers initially draft a research proposal, detailing the planned research protocol and writing the introduction and methodology sections in the future tense. Before data collection begins, the manuscript is submitted to reviewers, who may request revisions. After addressing these requests and satisfying both reviewers and editor, Stage 1 concludes with the manuscript's provisional acceptance. Subsequently, researchers proceed to collect, analyze, and report data as pledged in Stage 1. Notably, a proposal accepted in Stage 1 is guaranteed publication in Stage 2, regardless of the data outcomes. This format aims to mitigate the bias towards publishing positive findings (e.g., publication bias), which can skew the scientific record.

Innovating Journal Formats to Accelerate Psychological Science

The registered report format enables numerous creative developments. For instance, peer-reviewed journals can innovate in their approach to handling reviews, and platforms like *Routledge Open Research* have already begun exploring non-traditional peer-review processes.

Imagine a scenario where, upon article acceptance, an announcement is made offering authorship titled “Replicators” to those who conduct a preregistered replication under specified conditions. A similar approach is used in *Behavioral and Brain Sciences*, where a target article is first written and subsequently complemented by collected commentaries which the original authors then address; a new journal format could be introduced to enhance replication efforts. After the original study is accepted, replicators could be allotted a specific timeframe to conduct and report their replication studies. These replication reports could subsequently be published alongside the original study, akin to commentaries. Following this, a meta-analysis of the findings could be conducted. This comprehensive approach could accelerate the advancement of psychological science by leveraging a journal system that encourages thorough examination and replication of research findings.

Such advancements could significantly accelerate the progress of psychology. As a result, the introduction and discussion sections of empirical papers will resemble those in biology—concise and focused directly on the effects and their boundary conditions, rather than constructing elaborate narratives. Replications could further delineate these conditions, and a meta-analytic evidence base built on preregistered findings could emerge, synthesizing disparate sources and propelling scientific advancement in a mutually beneficial system.

Theoretical papers play a crucial role in refining and testing psychological theories. They can propose specific empirical tests linked directly to a theory, clarify the inferences drawn from the theory, and continuously update theoretical expectations as new evidence emerges. For instance, in a collaborative effort, the expectations of a theory could be meticulously detailed item by item. The proponents of the theory could then refine these expectations during the peer review process, after which independent researchers could empirically test the most significant theoretical expectations. This process allows theories to evolve in a cumulative manner, similar to methodologies in the natural sciences.

Theory development is not commonly prioritized as a career focus among psychologists. However, promoting such a goal could significantly mature psychological theories through rigorous theoretical articles. These articles could also suggest methods for the mathematical modeling of empirical findings and assess the alignment of these models with empirical data. Recently, advancements in machine learning have made it possible to predict the likelihood of a study's replication even before its publication; however, such examples remain scarce in social sciences (Altmejd et al., 2019).



A robust theoretical framework should ideally specify from the outset the physical conditions (e.g., temperature between 20 and 30 degrees Celsius) and the specific cultural contexts and conditions (e.g., populations experiencing moderate scarcity, with adequate literacy and English proficiency) under which studies can be replicated. Currently, psychology lacks a comprehensive perspective that integrates all these variables effectively. This deficiency leads to another facet of the replication crisis: the theory crisis.

Theory Crisis

There is a notable absence of an inclusive and comprehensive theoretical model in psychology and the social sciences that can comprehensively explain human behavior (Gigerenzer, 2010; Meehl, 1967; Muthukrishna & Henrich, 2019). This deficiency stems from our inability to perceive the full spectrum of factors that predict human behavior, akin to the approaches used in the natural sciences. Additionally, a lack of genuine interdisciplinary collaboration and the isolated nature of research within fields exacerbate this problem.

A comprehensive theory should make precise predictions from the outset, significantly narrowing down the network of possibilities. Comprehensive theoretical models, equipped with defined assumptions about human behavior, should predict specific behavioral outcomes from the interplay of genetics, cognition, and environment. For example, if a theory, rooted in evolutionary logic or cultural evolutionary models, predicts that certain moral judgments (like loyalty, authority, sanctity) are more prevalent in contexts or cultures with high pathogen prevalence, and identifies cultural and individual factors shaping these judgments, it approaches a comprehensive model. However, the robustness of most effects remains uncertain, and there is limited data on the precise conditions under which they occur, hindering the development of comprehensive theoretical models.

Although existing theoretical models are not comprehensive, hypotheses derived from these models are often tested using deductive methods as though the underlying theories were robust. However, particularly in nascent disciplines where knowledge is still evolving, abduction—rather than deduction—can be a better alternative (Haig, 2018). This method, used by Darwin in developing natural selection, involves identifying solid facts first, which are subsequently explained through proto-theories and eventually modeled mathematically. Theory development should proceed by comparing these emerging theories against alternatives (Borsboom et al., 2021).

Thus, initiating theory development in psychology and the social sciences with merely verbal explanations and testing them through derived hypotheses might not be effective. Instead, robust effects and their boundary conditions should be identified using valid and reliable tools, and theories should be developed that best explain these findings. Historically, psychology has operated as if it were underpinned by robust theories, making deductive inferences. However, to develop sound theories, we first need solid findings, underpinned by valid and reliable experimental methodologies. This necessity highlights another critical aspect of the ongoing crisis in psychological research.

Validity Crisis

The lack of structured measurement tools and manipulations has been a longstanding issue in psychology, impacting the reliability and validity of research findings (Cronbach & Meehl, 1955; Flake et al., 2022; Schimmack, 2021; Vazire et al., 2022). Problems with construct validity and ecological validity have been particularly noted across various studies (Vazire et al., 2022).

Furthermore, it is commonly observed that researchers often prioritize reporting reliability over validity, with scant attention given to validating findings (Flake et al., 2022). This oversight becomes a



significant concern in multi-laboratory and multicultural studies aimed at addressing the replication crisis. Typically, research tools validated in one cultural context are merely translated and used directly for hypothesis testing in other contexts without any preliminary validity checks.

A potential solution could involve distributing responsibilities among large groups of researchers, such as in studies with 100-200 authors, where half could focus on conducting validity studies while the other half carries out hypothesis testing. This approach would likely enhance the reliability of measurement and manipulation tools across different cultures. Without such measures, the clarity and accuracy of our scientific inferences are substantially compromised, leading us to another facet of the crisis in psychology.

Inference Crisis

What does a nonsignificant finding mean in a hypothesis test when the validity of the measurement tools is unknown? Without valid measurements, any test conducted is inherently unreliable, making it impossible to draw clear inferences. This issue is indicative of broader methodological challenges in behavioral and social sciences.

For instance, the lack of standardization in analysis and interpretation practices can lead to inconsistencies. Given that statistical analysis assumptions and data characteristics vary significantly, achieving consensus is challenging, and diverse traditions have emerged across fields. Some of these traditions have fostered robust experimental paradigms, while others have perpetuated statistical malpractices.

A common example is the practice of conducting mediation analysis on correlational data and using causal language, such as “mediation,” to describe findings derived solely from correlational methods. Although this approach was once prevalent in psychology, it has now been largely discredited (Rohrer et al., 2022). Consider dividing a 20-item unidimensional measure into two sub-dimensions of 10 items each, which will naturally show strong correlations. Coding one as an independent variable and the other as a mediating variable often leads to significant relationships due to common variance, not because there is a true mediation effect. True mediation implies a causal mechanism whereby the mediating variable accounts for part of the effect of the independent variable on the dependent variable. Such claims require a series of experiments: one where the independent variable is manipulated and the mediating variable is measured; another where the independent variable's effect on the dependent variable is examined; a third where the mediating variable is manipulated to observe its effect on the dependent variable; and finally, a study that manipulates the independent variable and measures both the mediating and dependent variables in a counterbalanced design (e.g., Bullock et al., 2010; Stone-Romero and Rosopa, 2011). Any simpler tests of mediation may not provide genuine causal insights, instead reflecting statistical artifacts.

Another questionable practice is the exclusion of participants who fail manipulation checks from analyses, arguing that they were not influenced by the manipulation. This approach, once common, can compromise the validity of an experiment by disrupting the random assignment (Varaine, 2023). Although widely used in the past, this method is now recognized as problematic, even among those who once applied it (see Saribay et al., 2020, p. 195).

Intention-Behavior Gap

A significant factor exacerbating the inference crisis in psychological research is the predominant reliance on intention-based self-report measurements, with actual behavior measurements becoming increasingly rare. Historically, classical experiments in social psychology primarily observed real



behaviors in laboratory settings or field experiments. However, since the 2000s, the convenience and practicality of intention-based self-report measures have dominated the field, despite their known limitations.

While fields like behavioral economics continue to uphold actual behavior measurement as a standard—actively avoiding reliance solely on intention measurements—the majority of psychological research has developed literatures based predominantly on intentions. The disconnect between what people intend and what they actually do has been well-documented and remains a critical concern (Sheeran & Webb, 2016; Yilmaz, 2022). This is particularly pronounced in moral psychology, where discrepancies between stated and actual moral behaviors are often influenced by factors like social desirability (Bostyn et al., 2018, Bostyn et al., 2019).

In behavioral economics, moral judgments, often assessed through intention-based measures like the Moral Foundations Questionnaire, are explored using incentivized economic games. This approach highlights the potential benefits of integrating actual behavior measurements into research methodologies. As such, alongside the continued use of intention measures, it is imperative for the advancement of the field that greater emphasis be placed on developing and utilizing methodologies that capture actual behaviors.

WEIRD Crisis

Henrich et al. (2010) highlight a significant limitation in psychological research: the overwhelming majority of findings are derived from studies involving US or European university students, a group that represents only a narrow slice of global cultural diversity. This cohort, characterized as WEIRD, exhibits specific cultural traits that are not necessarily representative of the global population. Henrich (2021) points out that behaviors from WEIRD populations often deviate significantly from those of the general global populace, emphasizing the peculiarity of relying so heavily on this group for psychological insights.

The necessity of verifying whether psychological findings hold true across different cultures is crucial for theoretical progress. Although contemporary multicultural studies aim to address this issue, the continued reliance on university students and online platforms for data collection complicates these efforts. Moreover, while modern technology facilitates access to data from diverse cultural backgrounds, truly testing our findings and drawing comprehensive conclusions across varied demographic groups, such as the elderly or hunter-gatherers who remain underrepresented in research, hinges on resolving the crisis of representation.

Strategies to Overcome Fundamental Challenges in Psychological Sciences

The issues delineated in this paper underscore the foundational challenges that behavioral sciences face, with no simple solutions readily available. However, it has become evident which practices are ineffective and may exacerbate these issues. In this concluding section, I will outline practical strategies that researchers can adopt to navigate these challenges effectively.

Applying the Abductive Method

Similar to a detective gathering all possible evidence to solve a case, or a physician collecting data to diagnose a disease, researchers in psychology can benefit from the abductive method of reasoning to develop theories. This approach involves meticulously collecting empirical data, testing various explanations against this data, and determining the most plausible explanation. In psychology,



where comprehensive theoretical frameworks are often lacking, viewing existing theories as tentative and focusing on establishing causal relationships between key variables is crucial. Researchers should start by understanding these relationships, formulating verbal explanations, modeling them mathematically, and then progressing to formal theory development. The Abductive Theory of Method (ATOM) provides a structured approach for applying abductive reasoning in psychological research, offering a comprehensive model for theory development from the ground up (Borsboom et al., 2021; Haig, 2018; Muthukrishna & Henrich, 2019).

Prioritizing Validity Studies

In psychology, the validity of measurements—beyond straightforward metrics such as age or gender—is often compromised by the complexity of the subject matter and questionable research practices. The importance of validating measurement instruments is frequently overlooked (Schimmack, 2021), which undermines the reliability of research findings, particularly in replication studies involving multiple laboratories. A fundamental shift in practice may involve assigning half of a large group of researchers (in projects with 100-200 authors) to conduct validity studies while the other half performs hypothesis testing with tools validated through factor analysis. This division of labor is not common practice; however, it can significantly enhance the integrity and credibility of psychological research.

Recognizing the Limits of Inferences

Just as we should not infer causality from a correlational study, we should be very careful about using causal language such as “mediation effect”. Likewise, we should adhere to open science practices, particularly in statistical analyses, pre-registering the entire research protocol and clearly distinguishing between confirmatory and exploratory hypotheses and shaping our conclusions accordingly. We should determine the boundary conditions of our findings and conclude the research by clearly considering under which conditions it can be replicated. Similarly, when removing participants from the study according to certain criteria, we should be aware of the rationale for this and proceed with a systematic logic.

Increasing Representativeness

Another crucial consideration when delineating the boundaries of our conclusions is recognizing the cultural and contextual conditions in which the research is being conducted and acknowledging the necessity of replicating the effect under different conditions. For example, a psychologist working in moral psychology should follow anthropological or animal studies that examine moral sensitivities in animals or that are based on actual behavior in hunter-gatherer societies and try to integrate the methodological and theoretical diversity of adjacent fields such as behavioral economics. When a finding is made with a sample of students, it is of great importance for the progress of the field to re-test this finding in a more representative sample and in different large data sets including different cultures in order to develop a holistic picture and knowledge about the boundary conditions.

Creating Behavioral Measures

While social psychology, in particular, appears to be moving away from behavioral measures and hence away from the influence of context, in fact the development of the field has mostly been through behavioral experiments that take context seriously. Therefore, it is crucial to make an effort to incorporate big data-based human behavior into research, such as the economics games used in fields like behavioral economics and social media posts or Google Trends. Additionally, greater emphasis should be placed on developing original behavioral tasks that can measure the concepts we use.

Building Big Teams

It is highly challenging for a researcher to accomplish all of this alone. This is why science is one of the greatest forms of collaboration ever developed by humankind, and why working in large teams is essential for achieving these objectives. The MINT Lab (www.moralintuitionslab.com), which I lead in Türkiye, is an example of such an organization as an umbrella laboratory that includes social psychologists, behavioral economists, data scientists, graduate and undergraduate students.

Conclusion

Currently, the landscape of behavioral sciences reveals several critical challenges. Most findings in this field exhibit small effect sizes, and the lack of a comprehensive theoretical model constrains our ability to make reliable policy recommendations. Additionally, it is important to recognize that much of our knowledge is derived from WEIRD populations, which may not represent global human behavior. This, coupled with significant validity issues in our measurement tools and a lack of standardization in our inferential processes, suggests that our findings might sometimes contribute to spreading questionable inferences throughout the discipline.

Given that even established fields like physics are experiencing theoretical crises, it is neither surprising nor disheartening that psychology, a relatively young discipline, is still in its early stages of development. The open science movement has crucially helped us to accurately identify these issues and has provided a roadmap for gradual but essential progress. In this article, I have offered several reflections and recommendations for the future of behavioral sciences, drawing from my personal experiences and subjective insights.

Looking ahead, it will be intriguing to see how psychology evolves in the era of artificial intelligence over the next decade. The developments in this field and their alignment with the perspectives shared here are something we can eagerly anticipate and observe.

List of Basic Texts and Further Readings

Borsboom, D., van der Maas, H. L., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory construction methodology: A practical framework for building theories in psychology. *Perspectives on Psychological Science*, 16(4), 756-766. <https://doi.org/10.1177/1745691620969647>

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281-302. <https://doi.org/10.1037/h0040957>

Flake, J. K., Davidson, I. J., Wong, O., & Pek, J. (2022). Construct validity and the validity of replication studies: A systematic review. *American Psychologist*, 77(4), 576-588. <https://doi.org/10.1037/amp0001006>

Gigerenzer, G. (2010). Personal reflections on theory and psychology. *Theory & Psychology*, 20(6), 733-743. <https://doi.org/10.1177/0959354310378184>

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?. *Behavioral and Brain Sciences*, 33(2-3), 61-83. <https://doi.org/10.1017/S0140525X0999152X>

Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science*, 34(2), 103-115. <https://doi.org/10.1086/288135>

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





Author Details Onurcan Yılmaz (Assoc. Prof.)

¹ Kadir Has University, Faculty of Economics, Administrative and Social Science, Psychology, Istanbul, Türkiye

0000-0002-6094-7162

onurcan.yilmaz@khas.edu.tr

References

- Altmejd, A., Dreber, A., Forsell, E., Huber, J., Imai, T., Johannesson, M., Kirchler, M., Nave, G., & Camerer, C. (2019). Predicting the replicability of social science lab experiments. *PLoS ONE*, 14(12), e0225826. <https://doi.org/10.1371/journal.pone.0225826>
- Bench, S. W., Rivera, G. N., Schlegel, R. J., Hicks, J. A., & Lench, H. C. (2017). Does expertise matter in replication? An examination of the reproducibility project: Psychology. *Journal of Experimental Social Psychology*, 68, 181-184. <https://doi.org/10.1016/j.jesp.2016.07.003>
- Borsboom, D., van der Maas, H. L., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory construction methodology: A practical framework for building theories in psychology. *Perspectives on Psychological Science*, 16(4), 756-766. <https://doi.org/10.1177/1745691620969647>
- Bostyn, D. H., Sevenhant, S., & Roets, A. (2018). Of mice, men, and trolleys: Hypothetical judgment versus real-life behavior in trolley-style moral dilemmas. *Psychological Science*, 29(7), 1084-1093. <https://doi.org/10.1177/0956797617752640>
- Bostyn, D. H., Sevenhant, S., & Roets, A. (2019). Beyond physical harm: How preference for consequentialism and primary psychopathy relate to decisions on a monetary trolley dilemma. *Thinking & Reasoning*, 25(2), 192-206. <https://doi.org/10.1080/13546783.2018.1497536>
- Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what's the mechanism? (don't expect an easy answer). *Journal of Personality and Social Psychology*, 98(4), 550-558. <https://doi.org/10.1037/a0018933>
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281-302. <https://doi.org/10.1037/h0040957>
- Curry, O. S., Mullins, D. A., & Whitehouse, H. (2019). Is it good to cooperate? Testing the theory of morality-as-cooperation in 60 societies. *Current Anthropology*, 60(1), 47-69. <https://doi.org/10.1086/701478>
- Errington, T. M., Mathur, M., Soderberg, C. K., Denis, A., Perfito, N., Iorns, E., & Nosek, B. A. (2021). Investigating the replicability of preclinical cancer biology. *Elife*, 10, e71601. <https://doi.org/10.7554/eLife.71601>
- Flake, J. K., Davidson, I. J., Wong, O., & Pek, J. (2022). Construct validity and the validity of replication studies: A systematic review. *American Psychologist*, 77(4), 576-588. <https://doi.org/10.1037/amp0001006>
- Gigerenzer, G. (1991). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review*, 98(2), 254-267. <https://doi.org/10.1037/0033-295X.98.2.254>
- Gigerenzer, G. (2010). Personal reflections on theory and psychology. *Theory & Psychology*, 20(6), 733-743. <https://doi.org/10.1177/0959354310378184>
- Götz, F. M., Gosling, S. D., & Rentfrow, P. J. (2022). Small effects: The indispensable foundation for a cumulative psychological science. *Perspectives on Psychological Science*, 17(1), 205-215. <https://doi.org/10.1177/1745691620984483>
- Haig, B. D. (2018). An abductive theory of scientific method. *Method Matters in Psychology: Essays in Applied Philosophy of Science*, 35-64. https://doi.org/10.1007/978-3-030-01051-5_3
- Haig, B. D. (2019). The importance of scientific method for psychological science. *Psychology, Crime & Law*, 25(6), 527-541. <https://doi.org/10.1080/1068316X.2018.1557181>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?. *Behavioral and Brain Sciences*, 33(2-3), 61-83. <https://doi.org/10.1017/S0140525X0999152X>
- Henrich, J. (2021). *The WEIRD people in the world: How the west became psychologically peculiar and particularly prosperous*. Penguin Books.



- Ijzerman, H., Lewis Jr, N. A., Przybylski, A. K., Weinstein, N., DeBruine, L., Ritchie, S. J., Vazire, S., Forscher, P. S., Morey, R. D., Ivory, J. D., & Anvari, F. (2020). Use caution when applying behavioural science to policy. *Nature Human Behavior*, 4(11), 1092-1094. <https://doi.org/10.1038/s41562-020-00990-w>
- Isler, O., Yılmaz, O., & Dulleck, U. (2023). Scarcity improves economic valuations when cognitively salient. *Journal of Economic Behavior & Organization*, 209, 362-371. <https://doi.org/10.1016/j.jebo.2023.02.019>
- Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does "failure to replicate" really mean?. *American Psychologist*, 70(6), 487-498. <https://doi.org/10.1037/a0039400>
- Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science*, 34(2), 103-115. <https://doi.org/10.1086/288135>
- Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behavior*, 3(3), 221-229. <https://doi.org/10.1038/s41562-018-0522-1>
- Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Struhl, M. K., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. N., Scherer, L. D. Schönbrodt, F. D. & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, 73, 719-748. <https://doi.org/10.1146/annurev-psych-020821-114157>
- Rohrer, J. M., Hünemund, P., Arslan, R. C., & Elson, M. (2022). That's a lot to process! Pitfalls of popular path models. *Advances in Methods and Practices in Psychological Science*, 5(2), 25152459221095827. <https://doi.org/10.1177/25152459221095827>
- Saribay, S. A., Yılmaz, O., & Körpe, G. G. (2020). Does intuitive mindset influence belief in God? A registered replication of Shenhav, Rand and Greene (2012). *Judgment and Decision Making*, 15(2), 193-202. <https://doi.org/10.1017/S1930297500007348>
- Schimmack, U. (2021). The validation crisis in psychology. *Meta-Psychology*, 5. <https://doi.org/10.15626/MP.2019.1645>
- Shah, A. K., Shafir, E., & Mullainathan, S. (2015). Scarcity frames value. *Psychological Science*, 26(4), 402-412. <https://doi.org/10.1177/0956797614563958>
- Sheeran, P., & Webb, T. L. (2016). The intention-behavior gap. *Social and Personality Psychology Compass*, 10(9), 503-518. <https://doi.org/10.1111/spc3.12265>
- Stone-Romero, E. F., & Rosopa, P. J. (2011). Experimental tests of mediation models: Prospects, problems, and some solutions. *Organizational Research Methods*, 14(4), 631-646. <https://doi.org/10.1177/1094428110372673>
- Thaler, R. (1985). Mental accounting and consumer choice. *Marketing Science*, 4(3), 199-214. <https://doi.org/10.1287/mksc.4.3.199>
- Varaine, S. (2023). How dropping subjects who failed manipulation checks can bias your results: An illustrative case. *Journal of Experimental Political Science*, 10(2), 299-305. <https://doi.org/10.1017/XPS.2022.28>
- Vazire, S., Schiavone, S. R., & Bottesini, J. G. (2022). Credibility beyond replicability: Improving the four validities in psychological science. *Current Directions in Psychological Science*, 31(2), 162-168. <https://doi.org/10.1177/09637214211067779>
- Yılmaz, O. (2022). *The new genealogy of morality: A psychological and evolutionary perspective [Ahlakın yeni soyağacı: Psikolojik ve evrimsel bir bakış]*. Koç University Press.

