

This paper makes a timely and important contribution to the scientific foundations of cybersecurity by addressing a long-standing and consequential gap: the lack of a clear, operational understanding of reproducibility and replicability in security research. While these concepts are widely invoked across the field, their definitions remain inconsistent, their application uneven, and their relationship to scientific validity poorly understood. This challenge has been highlighted in influential efforts such as the National Academies of Sciences report on reproducibility and replicability, which clarifies definitions but stops short of providing an operational framework for applying them in practice. This work directly confronts that ambiguity and provides a unifying framework that advances cybersecurity as a rigorous scientific discipline.

The authors introduce the Tree of Validity (ToV), a novel and practical model that represents experimental studies as structured combinations of key elements: problem, domain, method, data, and analysis. By modeling these elements as a binary tree of “same” versus “different” choices, the ToV captures the full space of experimental variations. This enables precise characterization of how one study relates to another, moving beyond labels such as “reproduced” or “replicated” to a more nuanced understanding of experimental relationships. In doing so, the paper resolves a core tension in prior work, which has struggled to reconcile replicability as both a property of artifacts and an action undertaken by researchers.

Importantly, the ToV is not merely a conceptual taxonomy. It provides an operational lens for reasoning about validity, distinguishing between the potential for replication (what experiments are possible given available artifacts), the execution of a specific experimental path, and the conclusions drawn from that execution. This separation is particularly valuable in cybersecurity, where experiments are often constrained by incomplete artifact availability, ethical considerations, and rapidly evolving environments. The framework makes these constraints explicit and allows researchers to reason systematically about their impact on scientific claims.

The paper also demonstrates the applicability of the ToV across diverse areas of security research, including systems vulnerabilities, machine learning defenses, and large-scale measurement studies. These case studies highlight a critical insight: even when artifacts are available and experiments are rerun, differences in context can lead to divergent outcomes. The ToV provides a structured way to represent and communicate these differences, enabling more meaningful comparison of results and a clearer understanding of generalizability.

This work aligns strongly with the goals of the Science of Security initiative by advancing rigor, clarity, and methodological consistency in cybersecurity research. It complements

ongoing community efforts in artifact evaluation, reproducibility, and research infrastructure by offering a unifying framework that can be applied across domains and methodologies. In doing so, it helps bridge the gap between individual experimental practices and broader scientific reasoning about evidence and validity.

Beyond its immediate contributions, this paper has the potential to shape how cybersecurity research is conducted, evaluated, and communicated. By reframing reproducibility and replicability as a structured space of experimental possibilities rather than binary outcomes, it provides a foundation for more systematic, transparent, and cumulative scientific progress.