

The Power of Theory

J. GREGORY TRAFTON, Naval Research Laboratory

PAULA RAYMOND, Toxcel

SANGEET KHEMLANI, Naval Research Laboratory

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; • **Computer systems organization** → **Robotics**;

Additional Key Words and Phrases: Human robot interaction, psychology

ACM Reference format:

J. Gregory Trafton, Paula Raymond, and Sangeet Khemlani. 2021. The Power of Theory. *ACM Trans. Hum.-Robot Interact.* 10, 1, Article 11 (February 2021), 2 pages.

<https://doi.org/10.1145/3439716>

Hoffman & Zhao’s paper “A Primer for Conducting Experiments in Human-Robot Interaction” is an excellent paper overall with many practical suggestions throughout. We expect it to be used as a source for how to run human-subject experiments in the field of human-robot interaction, especially for people not trained directly in psychology.

If we want to advance our understanding of human-robot interaction, however, we must look beyond individual studies. As with any scientific endeavor, broad progress is driven by theory. In this reply, we will expand upon the importance of theory and the generation of hypotheses.

A theory is an overarching idea that explains existing data and generates predictions that can be tested. It must be consistent with what we know already, and we must be able to logically follow through to the implied consequences. In fact, if the predictions are contradicted by empirical study, the theory must be adjusted or discarded. Theories increase our understanding of the world not just by explaining what we already know, but also by establishing principles that predict testable outcomes.

Without a clear articulation of a theory, experimenters can find themselves with uninterpretable results. In HRI, two types of theories are most relevant: theories about artifacts (which we call **design theories**) and theories about people (which we call **psychological theories**). The difference between the two can be subtle.

Design theories provide some reason to believe that one robot or component functions differently than another.¹ For example, a researcher may create a new design pattern for completing an interaction (e.g., saying “goodbye”), believing that people will feel more satisfied about the

¹Improved algorithms, interaction styles, systems, and artifacts can all be considered design theory for this discussion

This work was supported by the U.S. Office of Naval Research and the Office of the Secretary of Defense to GT and SK. Authors’ addresses: G. Trafton and S. Khemlani, NRL, Code 5515, Washington, DC, 20375; emails: {greg.trafton, sunny.khemlani}@nrl.navy.mil; P. Raymond, toXcel, LLC, 7140 Heritage Village Plaza, Gainesville, Virginia 20155 USA; email: Paula.Raymond@toxcel.com.

This paper is authored by an employee(s) of the United States Government and is in the public domain. Non-exclusive copying or redistribution is allowed, provided that the article citation is given and the authors and agency are clearly identified as its source.

2021. 2573-9522/2021/02-ART11 \$15.00

<https://doi.org/10.1145/3439716>

interaction with an explicit finale. The hypothesis should follow from the theory: that people feel more satisfied with their interaction when the design pattern is used compared to alternative design patterns, or no design pattern at all. Determining exactly how to measure “more satisfied” will likely overlap with other theories or constructs from other fields, and it should certainly use excellent experimental design, but note that the primary theoretical work occurs at the design stage, not at the experimental stage. Also note that the design pattern itself may have implications for psychological theory, but it does not have to.

A **psychological theory** in HRI makes claims about how human cognition and behavior can change given different interactions. For instance, one psychological theory might posit that memory might be less disrupted at specific conversational breakpoints or task boundaries than at others. The hypothesis should again follow directly from the theory: robots that interrupt a teammate at task boundaries during a conversation may help people remember more than a robot that interrupts a person at an inappropriate time. An experiment designed to test interruption behavior in robots helps investigators uncover truths about human psychology along with figuring out how to build better robots.

When evaluating a design theory, the theoretical effort typically occurs before the creation of the experiment. The hypotheses generally ask the question of whether the new design is better than existing designs or no design at all. The experimental design must use sound methodology to accurately support the hypothesis, but the results of the experiment should provide evidence that the new design is more successful (or less) than existing designs. The results of the experiment may have implications for psychology or other fields.

When evaluating a psychological theory, the hypotheses are generated from the theory under consideration. Psychological theories are not an amalgam of ad hoc observations, nor are they an unrelated collection of hypotheses. Psychological theories have, at their core, a central unifying idea that serves to marshal and constrain their underlying principles. We appeal to psychological theories when we wish to understand why humans think, emote, remember, reason, socialize, or carry out other behaviors. The results of the experiment may (and hopefully will!) have implications for design or other fields.

Understanding how to generate hypotheses from either design theories or psychological theories is an important component of success for any empirical study performed in HRI.

ACKNOWLEDGMENTS

The views and conclusions contained in this document should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Navy. Thanks to Ben Wright for comments on a previous draft.