Closed-Loop Scientific Discovery in the Behavioral Sciences

Sebastian Musslick[1,2,3,*], Daniel Weinhardt[1], John Gerrard Holland[4], Younes Strittmatter[3,5]

[1] Institute of Cognitive Science, Osnabrück University, 49080 Osnabrück, Germany
[2] Department of Cognitive and Psychological Sciences, Brown University, Providence, RI 02906, USA
[3] Carney Institute for Brain Science, Brown University, Providence, RI 02906, USA
[4] Center for Computation and Visualization, Brown University, Providence, RI 02906, USA
[5] Department of Psychology, Princeton University, NJ 08544, USA
[*] Correspondence: sebastian.musslick@uos.de

Introduction to Discovery Problem. Behavioral sciences aim to elucidate the cognitive mechanisms underlying human behavior. Yet, the pace of behavioral research is constrained by the rate at which scientists can alternate between the design and execution of behavioral experiments, on the one hand, and the modeling of cognitive processes to explain human behavior on the other hand (Musslick et al., in press). To address these challenges, we introduce a framework for closed-loop scientific discovery in the behavioral sciences (Musslick, Strittmatter & Dubova, 2024). Specifically, we introduce Automated Research Assistant (AutoRA)—a system capable of automating and integrating steps of the behavioral research process, including experimental design, data collection, and model discovery (Musslick, Strittmatter & Holland, 2023). We showcase the capabilities of this system to autonomously discover novel behavioral phenomena and corresponding models of human cognition across three domains of cognitive psychology: perceptual decision-making, reinforcement learning, and cognitive control.

Overview of Closed-Loop System. AutoRA implements the autonomous empirical research paradigm, which involves a dynamic interplay between two artificial agents (Figure 1). The first agent, a *theorist* component, is primarily responsible for constructing computational models by relying on existing data to link experimental conditions to dependent measures. The second agent, an *experimentalist* component, is tasked with designing follow-up experiments that can refine and validate the models generated by the theorist. To close the loop for automated scientific discovery, both agents interface with *experiment runner* components that enable automated behavioral data collection from human participants via web-based experiments.

Components and Workflow. To illustrate the function and integration of each component within the AutoRA system, we consider an exemplary research study where a behavioral researcher aims to investigate how the probability of human participants detecting a coherent motion among a set of randomly moving dots varies as a function of the dot motion coherence.

The *experimentalist* components assume the responsibilities of research design experts, determining the subsequent iteration of behavioral experiments. These components utilize active learning techniques to generate a set of experimental conditions for experimental variables (e.g., a list of dot motion coherences to probe). To identify these conditions, experimentalists may take as input a set of candidate models provided by the theorist components, a record of previously tested conditions, and/or corresponding behavioral observations.

The *experiment runner* components function as research technicians, collecting behavioral data through web-based experiments (e.g., to assess participants' accuracy of detecting the correct motion as a function of dot motion coherence). Experiment runners typically accept the experimental conditions from experimentalists as input and produce behavioral observations as output. These runners interface with web servers for hosting experiments (e.g., Google Firebase), platforms for recruiting human participants (e.g., Prolific), and databases for storing the collected data (e.g., Google Firestore).

Finally, the *theorist* components embody the role of computational scientists, applying discovery algorithms to derive scientific models (e.g., statistical, mathematical, computational, or verbal) that best characterize the collected behavioral data. For instance, AutoRA interfaces with equation discovery methods to identify psychophysical laws that relate dot motion coherence to the probability of detecting the motion (Hewson, Strittmatter, Marinescu, Williams, & Musslick, 2023). Additionally, AutoRA includes theorists that combine recurrent network modeling with equation discovery to infer latent cognitive dynamics from noisy behavioral data (Weinhardt, Eckstein, & Musslick, 2024). Theorists take as input a set of experimental conditions and the corresponding behavioral observations obtained from experiment runners.

Generated results. AutoRA generated discoveries from real-world behavioral experiments across three distinct psychological paradigms. First, in the study of perceptual decision-making, the closed-loop system iteratively discovered an exponential law that describes how the ability to detect the motion of randomly moving

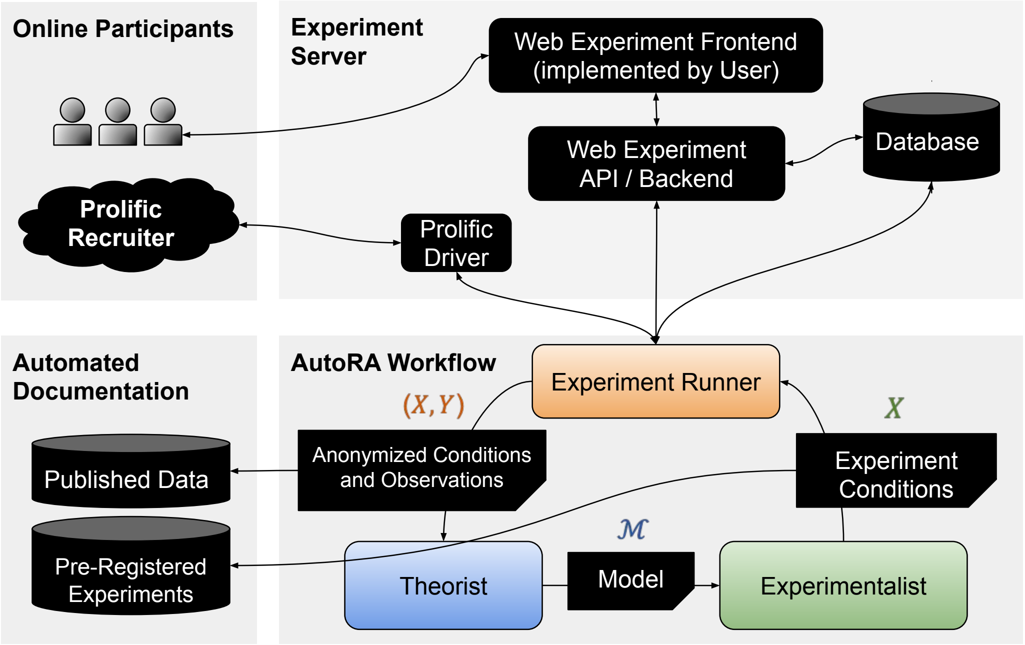

Figure 1: AutoRA framework for closed-loop behavioral research. AutoRA integrates components (colored boxes) to facilitate a closed-loop discovery process. Experiment runners (orange) automate data collection via online experiments and recruiting platforms (e.g., Prolific). Theorist components (blue) identify computational models $\mathcal{M}$ based on the data, while experimentalist components (green) suggest new experimental conditions X which are then executed to gather observations Y via experiment runner components. AutoRA also supports automated documentation of the research process and data via large language modeling.

dots improves as a function of the signal-to-noise ratio. In the domain of reinforcement learning, where participants chose between options with varying rewards, the system uncovered novel reinforcement learning rules that better explain human learning processes than traditional models. Specifically, it discovered a curiosity-based mechanism that steadily increases the value of non-chosen options. Lastly, when exploring a high-dimensional space of behavioral experiments involving task switching, AutoRA autonomously identified novel psychological phenomena. For instance, it discovered that humans tend to prefer completing tasks with lower difficulty before tackling more challenging tasks when required to execute them in rapid succession but not in slow succession. These results demonstrate that AutoRA possesses the capability to autonomously generate novel insights in the behavioral sciences, uncovering previously unobserved patterns and principles across domains of cognition.

References

Hewson, J. T. S., Strittmatter, Y., Marinescu, I., Williams, C. C., & Musslick, S. (2023). Bayesian Machine Scientist for Model Discovery in Psychology. In *NeurIPS 2023 AI for Science Workshop*.

Musslick, S., Bartlett, L. K., Chandramouli, S. H., Dubova, M., Gobet, F., Griffiths, T. L., Hullman, J., King, R. D., Kutz, J. N., Lucas, C. G., Mahesh, S., Pestilli, F., Sloman, S. J., & Holmes, W. R. (in press). Automating the Practice of Science: Opportunities, Challenges, and Implications. *Proceedings of the National Academy of Sciences*.

Musslick, S., Strittmatter, Y., & Dubova, M. (2024). Closed-loop computational discovery in the behavioral sciences. *PsyArXiv*. https://doi.org/10.31234/osf.io/c2ytb

Musslick, S., Strittmatter, Y., & Holland, J. G. (2023). AutoRA: Automated Research Assistant for Closed-Loop Computational Discovery. *Zenodo*. https://doi.org/10.5281/zenodo.10277415

Weinhardt, D., Eckstein, M. K., & Musslick, S. (2024). Computational discovery of human reinforcement learning dynamics from choice behavior. *NeurIPS 2024 Workshop on Behavioral ML*.