Workshop on Scaling Cognitive Science

Jordan W. Suchow (jws@stevens.edu)
Stevens Institute of Technology,
Hoboken, NJ 07030 USA

Thomas L. Griffiths (tomg@princeton.edu)
Princeton University,
Princeton, NJ 08540 USA

Joshua K. Hartshorne (joshua.hartshorne@bc.edu)
Boston College,
Chestnut Hill, MA 02467 USA

Keywords: web and mobile experiments, crowdsourcing, big data, large-scale experiments, citizen science

Significance

The proliferation of web-connected devices has presented significant opportunities and challenges to cognitive science — *opportunities* in that cognitive scientists can collect data relevant to human cognition orders of magnitude faster than before, addressing questions that were otherwise impossible to address; and *challenges* in that cognitive scientists require new infrastructure to collect these data and new methods to analyze them once collected.

This workshop brings together cognitive scientists who are at the forefront of these opportunities and challenges in scaling cognitive science, along with cognitive scientists who would like to be, to engage in a day of interactive exchange and development of ideas related to scaling cognitive science.

The workshop is a full day. Each presentation addresses a different opportunity or challenge. One set of presentations highlights opportunities. A second set highlights challenges of statistical analysis and data collection. In a break-out session, attendees address these opportunities and challenges head on.

Planned contributions of presenters

Below, we introduce the presenters and their planned contributions. Together, the presentations cover a wide variety of methods and application domains.

Joshua Hartshorne on Scaling Language Studies

Joshua Hartshorne is an assistant professor of psychology at Boston College, where he directs the Language Learning Laboratory. Hartshorne is interested in understanding what allows humans, but not current machines, to learn language — and why it is that children, despite their salient limitations of both cognition and experience, are so much more successful at language learning than adults. Hartshorne is particularly interested in applying new and emerging methods (such as computational modeling and crowdsourcing) to core problems in the language sciences.

Hartshorne will introduce the workshop by discussing opportunities and challenges associated with scaling cognitive science. Hartshorne will discuss these challenges in the context of running GamesWithWords.org, a web-based research laboratory that has tested hundreds of thousands of participants in a variety of experiments on language and cognition. Hartshorne will also discuss Pushkin, a platform for running large-scale online experiments.

Andrea Simenstad on Scaling the Zooniverse

Andrea Simenstad is a developer with the Zooniverse project at the University of Minnesota, where she collaborates with researchers, museum professionals, volunteers, and other developers to create crowdsourced citizen-science research projects with Zooniverse. Simenstad has a B.A. in cognitive science from Carleton College.

Simenstad will discuss Zooniverse. Zooniverse is the world's largest and most popular platform for crowdsourced citizen-science research, with more than 1,000,000 volunteers. The goal of the Zooniverse organization is to enable research that would not be possible, or practical, otherwise. Zooniverse research results in new discoveries, datasets useful to the wider research community, and many publications.

Todd Gureckis on Scaling Brain Data via Behavior

Todd Gureckis is a professor at New York University, where he runs the Cognition and Computation lab. The goal of the lab's research is to better understand the memory, learning, and decision processes which allow humans to carry out intelligent and adaptive behaviors.

Gureckis will discuss scaling brain data through behavior. One goal of computational cognitive neuroscience is to infer the relationship between latent variables in a cognitive model and particular features of brain data (e.g., fMRI or EEG). Typically, the precision of this relationship is constrained by the amount of brain data that can be collected (e.g., the number of subjects in a fMRI study). In this talk, Gureckis will leverage ideas from the statistics literature on inference with missing data and hierarchical modeling to explore how one can potentially learn more about a neuroimaging signal by collecting *only additional behavioral data*.

Lauren Rutter on Scaling via Mobile Applications

Lauren Rutter is a post-doctoral research fellow in the Laboratory for Brain and Cognitive Health Technology. Rutter studies basic cognitive and affective processes that influence the etiology and maintenance of internalizing psychopathology across the lifespan.

Rutter will discuss mobile applications for scaling mental-health assessment as well as large-scale citizen-science experiments. Rutter will demonstrate that digital technologies provide benefit to a variety of clinical populations and participants from the general public to generate scientific insights that would not have been possible with traditional models.

Tom Griffiths on Scaling Cognitive Science

Thomas Griffiths is a professor of psychology and computer science at Princeton University, where he leads the Computational Cognitive Science group. Griffiths' research focuses on developing mathematical models of higher-level cognition and understanding the formal principles that underlie our ability to solve the computational problems we face in everyday life.

Griffiths will introduce the workshop by discussing opportunities and challenges associated with scaling cognitive science. Griffiths will discuss how novel approaches to data collection and analysis — particularly "big data" — can change psychological research.

James Houghton on Scaling Social Cognition

James Houghton is a doctoral candidate at the MIT Sloan School of Management. He explores the mechanisms and outcomes of social contagion using a combination of data science and massive online experimentation, informed by rigorous simulation-based theory building.

Houghton will discuss the Empirica platform in the context of his research on social contagion. Social contagion research has for decades explored the effect of social network structure on the adoption of a single belief or practice. However, new theoretical models show that interaction between diffusants has a strong influence on their patterns of adoption. Houghton will present an experimental design called the 'detective game' for studying the diffusion of interacting beliefs in a controlled setting, explaining how an online laboratory experiment handles the confounds of pre-existing biases and meanings while cost-effectively achieving the scale necessary to draw statistical inferences.

Joshua Peterson on Scaling Choices and Categories

Josh Peterson is a postdoctoral fellow at Princeton University, where he studies the relationship between the representations learned by deep neural networks and those learned by people.

Peterson will discuss scaling research on decision making and categorization. Technological advances have made it possible to observe human behavior at an unprecedented scale, and the massive datasets that result provide two exciting new tools for cognitive modeling. First, such datasets serve as a test of the generalizability of our models and theories and therefore have the potential to improve them. Second, they offer the essential fuel for training machine learning models. Peterson will present case studies within two of the oldest modeling paradigms in cognitive science: categorization and decision making.

Mayank Agrawal on Scaling Moral Decisions

Mayank Agrawal is a doctoral student in psychology and neuroscience at Princeton University. Agrawal is interested in building formal computational models to understand how humans learn and make decisions.

Agrawal will discuss scaling studies of moral decision making. Standard methods of exploratory data analysis are vulnerable to noise in large datasets. To combat this problem, he proposes a methodology called *Scientific Regret Minimization*, which focuses on minimizing errors with respect to the data that are predictable. Agrawal applies this methodology to large datasets in the domains of moral reasoning and economic choice, and demonstrates how this approach helps build powerfully predictive (and interpretable) models and identify interesting new phenomena.

Jordan Suchow on Scaling Experiment Design

Jordan Suchow is an assistant professor at Stevens Institute of Technology, where he runs the Cognition Lab, which focuses on computational models of human learning, memory, and decision-making. As part of DARPA's Next Generation Social Science program, Suchow led a team to develop Dallinger, an online experimental platform for running behavioral and social science experiments at the scales necessary to understand emergent social phenomena.

Suchow will discuss Dallinger and associated techniques for experiment design at scale.

Platform Presentations

Platform presentations will include Dallinger, Empirica.ly, Pushkin, Zooniverse, and jsPsych.

Structure of the Workshop

The full-day workshop is structured as a series of 30-minute talks, short presentations on experiment platforms, and a breakout session on open questions to promote application of the ideas shared in the workshop.

Introductory talk (30 minutes)
Opportunities example #1 (30 minutes)
Opportunities example #2 (30 minutes)
Break
Opportunities example #3 (30 minutes)
Opportunities example #4 (30 minutes)
Opportunities example #5 (30 minutes)
Lunch
Analysis challenge #1 (30 minutes)
Analysis challenge #2 (30 minutes)
Analysis challenge #3 (30 minutes)
Break
Platform advertisements (45 minutes)
Breakout sessions on open questions (45 minutes)