Experimental Behavioral Research for Designing Human Computational Systems

Andrew Mao
Harvard University
mao@seas.harvard.edu

Abstract

The growing complexity of human computation systems underlies a need for more accurate and realistic studies of human behavior in computer science. At the same time, the development of online experimental research is a valuable opportunity for computer scientists to study human behavior in a principled and causal way. I propose several ways in which novel experimental approaches combined with other computer science tools can advance the state of the art in human computation to build smarter systems that leverage individual and collective human behavior, and also answer more general questions in social science.

Introduction

Computer science has outgrown its hardware and algorithmic roots to encompass a broad variety of systems with human participation. Although early examples of this—peer-to-peer networks, Wikipedia, and display advertising—arose in a largely ad hoc manner, the recent design of such systems has become more deliberate and principled. Hybrid human-computer systems solve problems that neither can do alone (Khatib et al. 2011). Artificial intelligence is applied to the optimization of online crowd workers (Kamar, Hacker, and Horvitz 2012). The field of algorithmic game theory has adopted tools from economics for modeling and designing advertising auctions, online markets, and other Internet systems. Machine learning techniques have shown great promise in discovering patterns in real-world, human-generated social and textual data (Blei 2012).

A largely missing piece from this realm of research is a methodology for producing causal statements about the behavior of human agents under different conditions, especially in a realistic and generalizable way. Many examples of Internet systems were designed for the participation of large numbers of human participants through a great deal of trial-and-error, and it is often unclear whether the resulting collective behavior is as intended. Moreover, using data mining techniques to deduce causal patterns from observational data, even *big* data, is challenging or impossible in many circumstances. As a result, we often need the *right* data, and not just a lot of data, to answer a particular question of interest.

For decades, and across a wide variety of disciplines, the gold standard for determining causality has been the controlled, randomized experiment. While this has traditionally involved participants that are students in physical university labs, the Internet has quickly grown as a scalable source of experimental data. Software companies such as Google and Microsoft routinely conduct randomized experiments on their deployed applications, collecting massive amounts of behavioral data. Many research disciplines have turned to systems like Amazon Mechanical Turk (MTurk) as tools to recruit large numbers of participants quickly and on short notice, and shown that online experiments can be just as reliable as those in physical labs (Horton, Rand, and Zeckhauser 2011; Mason and Suri 2012; Germine et al. 2012). Yet, the methodology for conducting online experiments has been developing for only several years, and there is still much untapped potential.

Conducting experimental behavioral research on the Internet presents several unique opportunities for computer scientists. First, experiments are a natural complement to theory regarding human agents, allowing for realistic tests of modeling assumptions and behavior toward incentives. Second, experiments can be used to empirically evaluate designs systems involving human participants in a principled way. Finally, experiments are an exceptional source of bespoke datasets for human behavior under controlled conditions, and experimental data can be especially powerful for developing behavioral models. Moreover from a practical perspective, computer science is arguably better equipped for software-driven experiments than any other field.

The overall goal of my research is to study the behavior of online users and design better human computational systems by using novel experimental methods combined with other computational tools. In particular, I focus on novel online experimental methodologies that are particularly promising for tackling research problems inaccessible to other methods. Experimental methods also serve as a bridge for building interdisciplinary connections between computer science and other research fields that are concerned with human behavior. In expanding the scope of online behavioral research, we can reach a better understanding of both collective and individual behavior—for the design of human computational systems, but also for eventually approaching social questions beyond simply those in human computation.

Motivation

Behavioral experiments are particularly promising in tandem with computer science, for three main reasons.

As a complement to theoretical research. In fields such as economics and psychology, a strong experimental tradition goes hand-in-hand with theoretical models of behavior. As a result, progress in theory research informs the design of experiments, and experimental results in turn promote more accurate theoretical models. For example, experiments and other empirical work in economics have resulted in a large literature of models accounting for limits to rational behavior (Gigerenzer and Selten 2002).

Computer science has developed deep connections to economics in the field of algorithmic game theory (Nisan et al. 2007), using the tools of mechanism design, auction theory, and game theory to model Internet-based systems such as online advertising, networks, peer production, and more recently crowdsourcing and human computation. However, there has been little experimental research supporting the burgeoning literature of theoretical work. In the recent 2014 ACM Conference on Economics and Computation, approximately only 10 out of 81 papers represented primarily empirical or experimental work, with only 4 papers consisting of new experiments—despite the specific recent addition of an empirical/experimental submission track.

When purely theoretical work proceeds independently of empirical verification, we risk pursuing poorly motivated problems or developing models that are far removed from reality. Hence, a strong experimental tradition can strengthen the interdisciplinary connections between computer science and social science, and also greatly improve the quality and relevance of theoretical research in computer science—just as it has in many other fields.

For principled design of human computational systems. There are many examples of Web-based communities or other online systems that have been developed to support large numbers of participants through a process of mainly trial-and-error—StackOverflow (Q&A), Reddit (content aggregation), Wikipedia (peer production), and GitHub (open-source software) are just a few examples. A contingent of computer science research, especially in the data mining community, has used this large amount of observational data to characterize the patterns in these communities and understand how they are successful. However, it is generally very difficult to derive causal claims from observational data without having statistically ideal instrumental variables to separate correlation from causation.

Although the aforementioned systems developed in an ad hoc fashion, online experiments now provide a principled tool for informing the design of such systems and optimizing participants' behavior. There are clear advantages in using online experiments to understand behavior web-based systems compared to studies using physical behavioral labs. While physical lab experiments must inevitably face questions of external validity by in generalizing to the real-world environments they simulate (Winer 1999), online experiments can be designed to be almost indistinguishable from the web-based environments they represent. Research in the areas of crowdsourcing, human computation, and human-computer interaction exemplifies this synergy.

For human computation in particular, experimental research is promising for several new avenues. Perhaps motivated by the concepts of divide-and-conquer in algorithms, a great deal of existing work focuses on how *workflows* can be used to decompose complex tasks into small and skill-agnostic tasks that can be done by any worker (Dai, Mausam, and Weld 2010; Kulkarni, Can, and Hartmann 2012). However, this may be an artifact of the status quo of microtask crowdsourcing, and dismisses benefits from humans being able to interact and coordinate, collaborate, or compete on different tasks. On the other hand, experiments in *collective intelligence* (Woolley et al. 2010) have shown that groups of humans can be quite collaborative, and may benefit from coordination mechanisms that enable them to work together and share knowledge. For example, systems such as Mobi and Cobi (Zhang et al. 2012; Kim et al. 2013) demonstrate how software-mediated collaboration can be used to tackle a complex task. While modeling behavior in these systems is difficult to approach directly with theory, experiments can be used to empirically compare one design over another, providing observational insights and eventually a basis for theoretical work.

As a source of controlled behavioral data. The field of *computational social science* (Lazer et al. 2009) has promised to revolutionize the way human behavior is studied by using data to draw inferences about and discover patterns in social systems. In parallel, the advent of *model-based machine learning* (Bishop 2013; Blei 2012), with its focus on descriptive rather than predictive models of data, has enabled a vast range of concise models for discovering patterns in social or textual data. Examples include inference of the underlying ideology of politicians (Gerrish and Blei 2012), the structure of e-mail communication networks (Krafft et al. 2012), and the implicit structure of social groups (Blundell, Beck, and Heller 2012).

However, when using data to study human behavior, more is not always better. Among other reasons, boyd and Crawford (2012) argue that large datasets are limited to what is available, may abstract away details available in smaller datasets, and may force one to inevitably discard some of the lower-level details along the way. Indeed, many of the particularly exciting descriptive models in computational social science have emerged from small or medium-sized datasets, or have limited tractability for very large amounts of data.

These applications are a perfect fit for experimental research, which can generate medium to large-sized datasets with behavior consisting of **only the data of interest** and **under different conditions**. Instead of drowning in observational data, experiments allow for a bespoke data set to be *designed* to answer the question at hand (Salganik 2014), under conditions that are controlled, and without the additional noise that must be filtered out for data mining. As a result, experimental data can combine with appropriate models to produce better insights than either approach alone.

Convergence to the Online Laboratory

Despite the many synergies that exist between behavioral experimental research and many aspects of computer science, experimental studies are still relatively uncommon as a methodology. A primary reason is that research involving human subjects has become prevalent only recently outside of human-computer interaction. Moreover, the principles of experimental design are not typically used in computer science research, and as a result experimental methodology is not well established in the field. One prominent exception, driven by the demands of large companies to test and optimize user interfaces, is that of large-scale, web-based field experiments in the data mining community (Kohavi et al. 2009; Tang et al. 2010; Bakshy, Eckles, and Bernstein 2014).

At the same time, online experiments are increasingly attractive to many disciplines, enjoying many advantages over their historical counterparts in the physical lab. Recruiting online subjects is more economical in scale, allowing for higher throughput and access to more participants as well as a greater diversity of participants than typical university labs. This in turn allows for faster iteration in designing experiments, collecting data, and designing new experiments, free of the constraints of scheduling undergraduate participants in a physical lab. Online experiments also promise greater external validity when designed for online tasks, and allow for experiment designs that were previously impossible with one hour of participation in a lab. For all these reasons, online experiments narrow the spectrum between highly controlled experiments and naturally occuring data (List 2008), and provide new ways to tackle interesting research questions—including those that have been previously difficult to approach.

A common platform for experimental research has been MTurk, established simultaneously as a methodology in many fields including psychology (Paolacci, Chandler, and Ipeirotis 2010; Buhrmester, Kwang, and Gosling 2011), political science (Berinsky, Huber, and Lenz 2012), and linguistics (Sprouse 2011). Mason and Suri (2012) pioneered methods for more complex behavioral experiments, describing techniques for recruiting subjects reliably and coordinating experiments with simultaneous multi-user participation. Paolacci, Chandler, and Ipeirotis (2010) studied the demographic composition of workers on MTurk, and Chandler, Mueller, and Paolacci (2013) showed that workers are much less naïve than many researchers imagine. As a result, online laboratories and MTurk in particular have become an apparatus for interdisciplinary behavioral research using methods and tools from many fields.

An Opportunity for Computer Science

With this rapid growth in online experimental research, there are many opportunities to use computer science techniques to tackle experimental problems, as well as questions of interest to computer science that would benefit from experimental techniques. A prime example is to study how users interact with each other in online systems and respond to different types of incentives. How can crowds self-organize to do complex tasks? Can we design crowd-powered computational systems that leverage interpersonal social behavior? How do we design optimal incentives for crowdsourcing and online communities? Experiments for such questions are of limited value when conducted in physical laboratories—in fact, distant relatives of today's web-based experiments were attempted in several decades ago in social science laboratories, but were eventually abandoned due to lack of progress (see Shure and Meeker (1970) for one of many examples). Modern web applications allow us to connect users together and instrument their behavior in a fine-grained way, all in a setting that portrays social online interaction much more accurately than any behavioral lab can reflect the physical world.

At the same time, there is much potential for innovation in the methodology of online human subject experiments. In deploying online experiments, researchers face a new challenge in engaging their participants and no longer have the luxury of a full hour of attention in a laboratory. Instructions that may have been previously delivered in a lengthy document or verbal presentation must now be concisely presented on the Internet, and participants need to be monitored to detect lapses in attention or even dropping out altogether. Deploying experiments thus involves elements of a distributed system for human agents, used in applications such as Legion (Lasecki et al. 2011) and in experimental frameworks such as TurkServer (Mao et al. 2012).

Current assumptions in recruiting paid participants from systems such as Amazon Mechanical Turk can be quite limiting. One obvious constraint is that experiments must take place over a short, contiguous period of time. By tracking participants' identities and storing some persistent information, we can enable asynchronous interactions between large numbers of people—hundreds, or even thousands—over longer periods of time, such as weeks or months. This methodology is possible even on MTurk itself, and allows for large-scale studies of interaction between networked users and or experiments of trading behavioral on financial or prediction markets. This upends the assertion of Zelditch (1969), who asked the rhetorical question of whether an "army" could be studied experimentally, concluding that it was both infeasible and unnecessary.

For studying crowd-powered systems, there are many examples of how experiments can scale beyond financial constraints on subject recruitment. The Zooniverse (Raddick et al. 2013; Reed et al. 2013), a citizen science platform, engages volunteer crowdsourcing participants to work on all manner of different scientific tasks, and is currently building infrastructre to enable real-time, large scale experimental behavioral studies in collaboration with researchers. At the same time, self-hosted online volunteer laboratories use large numbers of participants recruited without payment, but simply by providing something of value or interest. This approach has been used in studies of cognitive function (Halberda et al. 2012) and visual aesthetics (Reinecke and Gajos 2014) to produce studies with many thousands or even millions of participants and unprecedented amounts of data.

Proposed Research Directions

My primary interests are to use novel online experimental methodology combined with other computer science tools to study how human behavior and social interaction can be leveraged for more complex human computation and designing better online systems. My current and proposed research encompasses the following questions:

1. How do human users empirically respond to incentives in online systems, and how can we design these incentives to optimize their behavior?
2. Humans can self-organize for very complex tasks. How can we leverage collective intelligence and social interaction to develop smarter human computation systems?
3. How do we design models that more accurately capture how human agents respond to incentives, coordinate, and communicate?

I believe that improved methodology for online experiments are important for tackling these questions. First, they allow for **more realistic experiments**. In studying crowdsourcing and human computation, we can design online experiments that are very similar or even identical to their real-world applications, resulting more accurate behavioral observations and conclusions with a high degree of external validity. Second, we can construct **more interactive experiments** that capture fine-grained behavioral data and peer-to-peer interaction, allowing for a level of detail into social behavior that was unimaginable to the social science labs of 50 years ago. Finally, I propose that **better inference from experimental data** using descriptive modeling and model-based machine learning can combine with the data obtained in these realistic, interactive experimental studies to produce more insights into collective human behavior and principles for designing human computational systems.

Central to my research goal is to advance methodological boundaries in tandem with answering these questions, and as such I have been uniquely focused on developing tools and software to make experimental methods more accessible to the research community. By expanding on novel ways of recruiting human subjects for interactive experiments on the Internet (Mason and Suri 2012), I developed versions of Turkserver (Mao et al. 2012), now in its third iteration[1], an open-source framework that abstracts the common issues in building web-based experiments with real-time interaction between participants. TurkServer has enabled much of my research and has also allowed me to share accumulated expertise with other researchers and to deploy experiments in a scalable and reusable way. By promoting wider use and acceptance of experimental research using online participants, its development is crucial to my vision of behavioral research as an important part of computer science.

Preliminary Findings

Much recent crowdsourcing research has developed models for optimizing crowd work over different types of tasks, user abilities, and budgets, but there is little data on realistic behavior under the incentives from such models. In unpaid or volunteer crowdsourcing, such as the Zooniverse platform with its million registered users, workers are referred to as *citizen scientists* and contribute purely out of intrinsic motivation, absent of financial incentives. When considering incentives in crowdsourcing, a primary unresolved question is how the intrinsic incentives of volunteer workers relate to the financial ones of paid workers. I designed an experiment to compare workers paid by different piecewise financial incentives to volunteers on the same citizen science task (Mao et al. 2013). We found that paid workers could produce as good results as volunteers with proper controls, and could furthermore be implicitly induced to trade off between speed and quality by varying the scheme of payment.

One notable aspect of this experiment was the design of an interface for paid crowdsourcing workers almost identical to that used by volunteer citizen scientists, producing one of the first sets of data comparing intrinsically motivated and financially motivated workers on the same task. Using this data, and building on previous models of different worker abilities in classification tasks (Bachrach et al. 2012), I am designing a probabilistic model for tasks of detecting objects of interest in continuous data. This model allows for more nuanced measurement of possible biases from different conditions beyond a single measure of user ability, including false positive rate, false negative rate, and annotation precision. This model can capture richer details from the experimental data to develop realistic and more detailed models of how different financial incentives or intrinsic incentives produce biases in crowdsourced work.

More recently, I have used TurkServer to design an experiment to study how participants can self-organize in crisis mapping, a humanitarian phenomenon where many online participants collaborate in real time to generate reports of damage or casualties in response to a natural or man-made disaster. In contrast to many tasks typical of current online crowdsourcing, crisis mapping is not inherently parallelizable—users must coordinate and communicate with each other to share context while concurrently sifting through a large amount of data. Existing organization in crisis mapping has been ad hoc, and the number of natural disasters widely outpaces the availability of trained crisis mapping volunteers. An open area of research is how unspecialized, even anonymous, workers can to self-organize efficiently for complex tasks. Our experiment uses the crisis mapping task to produce records of real-time electronic interaction with the goal of developing both new theory and experimental methodology.

We have deployed the crisis mapping experiment with teams of online workers, and found that it allows for the control and manipulation power of a lab experiment while approaching the realism of a field experiment. Preliminary findings reveal different methods of collective self-organization, specialization in subgroups, spontaneous organization, and peer learning in complex tasks. Our study of crisis mapping is not only a novel example of *social* human computation, but is promising for answering fundamental questions about organizational behavior. As the theory of online, electronically mediated cooperation is only in its infancy, I am excited for how this project can inform both research in human computation and social science.

[1] https://github.com/HarvardEconCS/turkserver-meteor

References

Bachrach, Y.; Graepel, T.; Minka, T.; and Guiver, J. 2012. How to grade a test without knowing the answers—a bayesian graphical model for adaptive crowdsourcing and aptitude testing. *Proceedings of the 29th International Conference on Machine Learning (ICML).*

Bakshy, E.; Eckles, D.; and Bernstein, M. S. 2014. Designing and deploying online field experiments. In *Proceedings of the 23rd International World Wide Web Conference (WWW)*, 283–292. International World Wide Web Conferences Steering Committee.

Berinsky, A. J.; Huber, G. A.; and Lenz, G. S. 2012. Evaluating online labor markets for experimental research: Amazon. com's mechanical turk. *Political Analysis* 20(3):351–368.

Bishop, C. M. 2013. Model-based machine learning. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371(1984):20120222.

Blei, D. M. 2012. Probabilistic topic models. *Communications of the ACM* 55(4):77–84.

Blundell, C.; Beck, J.; and Heller, K. A. 2012. Modelling reciprocating relationships with hawkes processes. In *Proceedings of the 25th Annual Conference on Neural Information Processing Systems (NIPS)*, 2600–2608.

boyd, d., and Crawford, K. 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society* 15(5):662–679.

Buhrmester, M.; Kwang, T.; and Gosling, S. D. 2011. Amazon's mechanical turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science* 6(1):3–5.

Chandler, J.; Mueller, P.; and Paolacci, G. 2013. Nonnaïveté among amazon mechanical turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods.*

Dai, P.; Mausam; and Weld, D. S. 2010. Decision-theoretic control of crowd-sourced workflows. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI).*

Germine, L.; Nakayama, K.; Duchaine, B. C.; Chabris, C. F.; Chatterjee, G.; and Wilmer, J. B. 2012. Is the web as good as the lab? comparable performance from web and lab in cognitive/perceptual experiments. *Psychonomic bulletin & review* 19(5):847–857.

Gerrish, S., and Blei, D. M. 2012. How they vote: Issue-adjusted models of legislative behavior. In *Proceedings of the 25th Annual Conference on Neural Information Processing Systems (NIPS)*, 2753–2761.

Gigerenzer, G., and Selten, R. 2002. *Bounded rationality: The adaptive toolbox*. Mit Press.

Halberda, J.; Ly, R.; Wilmer, J. B.; Naiman, D. Q.; and Germine, L. 2012. Number sense across the lifespan as revealed by a massive internet-based sample. *Proceedings of the National Academy of Sciences* 109(28):11116–11120.

Horton, J. J.; Rand, D. G.; and Zeckhauser, R. J. 2011. The online laboratory: Conducting experiments in a real labor market. *Experimental Economics* 14(3):399–425.

Kamar, E.; Hacker, S.; and Horvitz, E. 2012. Combining human and machine intelligence in large-scale crowdsourcing. In *Proceedings of the 11th International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS).*

Khatib, F.; Cooper, S.; Tyka, M. D.; Xu, K.; Makedon, I.; Popović, Z.; Baker, D.; and Players, F. 2011. Algorithm discovery by protein folding game players. *Proceedings of the National Academy of Sciences* 108(47):18949–18953.

Kim, J.; Zhang, H.; Andr, P.; Chilton, L.; MacKay, W.; Beaudouin-Lafon, M.; Miller, R. C.; and Dow, S. P. 2013. Cobi: A community-informed conference scheduling tool. In *Proceedings of the 26th Symposium on User Interface Software and Technology (UIST).*

Kohavi, R.; Longbotham, R.; Sommerfield, D.; and Henne, R. M. 2009. Controlled experiments on the web: survey and practical guide. *Data Mining and Knowledge Discovery* 18(1):140–181.

Krafft, P.; Moore, J.; Desmarais, B.; and Wallach, H. M. 2012. Topic-partitioned multinetwork embeddings. In *Proceedings of the 25th Annual Conference on Neural Information Processing Systems (NIPS)*, 2807–2815.

Kulkarni, A.; Can, M.; and Hartmann, B. 2012. Collaboratively crowdsourcing workflows with turkomatic. In *Proceedings of the 15th ACM Conference on Computer Supported Cooperative Work (CSCW).*

Lasecki, W. S.; Murray, K. I.; White, S.; Miller, R. C.; and Bigham, J. P. 2011. Real-time crowd control of existing interfaces. In *Proceedings of the 24th Symposium on User Interface Software and Technology (UIST)*, 23–32. ACM.

Lazer, D.; Pentland, A. S.; Adamic, L.; Aral, S.; Barabasi, A. L.; Brewer, D.; Christakis, N.; Contractor, N.; Fowler, J.; Gutmann, M.; et al. 2009. Life in the network: the coming age of computational social science. *Science (New York, NY)* 323(5915):721.

List, J. A. 2008. Introduction to field experiments in economics with applications to the economics of charity. *Experimental Economics* 11(3):203–212.

Mao, A.; Chen, Y.; Gajos, K. Z.; Parkes, D.; Procaccia, A. D.; ; and Zhang, H. 2012. Turkserver: Enabling synchronous and longitudinal online experiments. In *Proceedings of the 4th Human Computation Workshop (HCOMP).*

Mao, A.; Kamar, E.; Chen, Y.; Horvitz, E.; Schwamb, M. E.; Lintott, C. J.; and Smith, A. M. 2013. Volunteering vs. work for pay: Incentives and tradeoffs in crowdsourcing. In *Proceedings of the 1st AAAI Conference on Crowdsourcing and Human Computation (HCOMP).*

Mason, W., and Suri, S. 2012. Conducting behavioral research on amazon's mechanical turk. *Behavior Research Methods* 44(1):1–23.

Nisan, N.; Roughgarden, T.; Tardos, E.; and Vazirani, V. V. 2007. *Algorithmic game theory*. Cambridge University Press.

Paolacci, G.; Chandler, J.; and Ipeirotis, P. G. 2010. Running experiments on amazon mechanical turk. *Judgment and Decision making* 5(5):411–419.

Raddick, M. J.; Bracey, G.; Gay, P. L.; Lintott, C. J.; Cardamone, C.; Murray, P.; Schawinski, K.; Szalay, A. S.; and Vandenberg, J. 2013. Galaxy Zoo: Motivations of citizen scientists. *Astronomy Education Review* 12(1):010106.

Reed, J.; Raddick, M. J.; Lardner, A.; and Carney, K. 2013. An exploratory factor analysis of motivations for participating in zooniverse, a collection of virtual citizen science projects. In *System Sciences (HICSS), 2013 46th Hawaii International Conference on*, 610–619. IEEE.

Reinecke, K., and Gajos, K. Z. 2014. Quantifying visual preferences around the world. In *Proceedings of the 32nd ACM Conference on Human Factors in Computing Systems (CHI)*, 11–20. ACM.

Salganik, M. J. 2014. (personal communication).

Shure, G. H., and Meeker, R. J. 1970. A computer-based experimental laboratory. *American Psychologist* 25(10):962.

Sprouse, J. 2011. A validation of amazon mechanical turk for the collection of acceptability judgments in linguistic theory. *Behavior research methods* 43(1):155–167.

Tang, D.; Agarwal, A.; O'Brien, D.; and Meyer, M. 2010. Overlapping experiment infrastructure: More, better, faster experimentation. In *Proceedings of the 16th International Conference on Knowledge Discovery and Data Mining (KDD)*, 17–26. ACM.

Winer, R. S. 1999. Experimentation in the 21st century: the importance of external validity. *Journal of the Academy of Marketing Science* 27(3):349–358.

Woolley, A. W.; Chabris, C. F.; Pentland, A.; Hashmi, N.; and Malone, T. W. 2010. Evidence for a collective intelligence factor in the performance of human groups. *science* 330(6004):686–688.

Zelditch, Jr., M. 1969. Can you really study an army in the laboratory? *A sociological reader on complex organizations* 528–39.

Zhang, H.; Law, E.; Miller, R.; Gajos, K.; Parkes, D.; and Horvitz, E. 2012. Human computation tasks with global constraints. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, 217–226. New York, NY, USA: ACM.