On Facilitating Human-Computer Interaction via Hybrid Intelligence Systems

WALTER S. LASECKI, University of Michigan – Ann Arbor

1. INTRODUCTION

What does the *perfect* interaction between a human and a computer look like? This question has driven much work in Human-Computer Interaction (HCI) over the past several decades, and while the answer (either explicitly or implicitly) typically ranges from "interaction should be natural" to "technology should be invisible" [Weiser 1999], there remains general consensus that a "socio-technical" gap will always remain a between the capabilities of systems, and the varied needs of people [Ackerman 2000].

This position paper suggests that Collective Intelligence (CI), in the form of Crowdsourcing for Human Computation (HCOMP), has the potential to bridge this socio-technical gap while enabling users to attain super-human performance on a wide range of tasks that they may seek to accomplish. In this paper, I outline possible intellectual directions for work in this space, and hope for this to act as a call-to-action for both crowdsourcing and interaction researchers.

2. BACKGROUND AND RELATED WORK

A significant amount of prior work in HCI, crowdsourcing, and human computation has led the field to this point. While it is a highly incomplete history of these fields, this section highlights a few key pieces of work to provide context for the discussion below.

2.1 Communication and Interaction

The way in which people have interacted with their environment and objects within it has been long studied (e.g., [Gibson 1977]). Without computational augmentation, mediation, or facilitation (provided by Social Computing Systems), these interactions were limited in scope. Conversations happened between smaller groups of people, over longer spans of time; physical gestures were only visible in face-to-face interactions; and the coordination of work was limited by the human time required to coordinate [Clark and Brennan 1991; Yates 1993].

With the advent of computer-mediated communication, these activities can be scaled, but not fundamentally changed [Olson and Olson 2000]. People can be said to interact with either `tools` or `agents` in *environments* (which provide an interactional context). Tools are artifacts used to enhance one's ability to complete a target task, and are used to directly accomplish a goal or subtask. Even complex tools are largely deterministic: when working properly, the outcome of an action taken with a tool—be it a hammer or a computer—is deterministic up to the stochasticity of the environment or context in which it is used. Agents, on the other hand, are entities that can reason or function on their own, and thus cannot be engaged with the deterministic assumptions we have when it comes to tools.

The importance of accurate mental models has been noted previously for both tools and agents [Bansal et al. 2019; Norman 2014]. Mental models are approximate understandings of operation that allow people to reason about interaction [Norman 2013]. While the need for mental models applies to agents and tools alike, the mental models people form of agents are not simply more complex versions of the ones we form for tools. Instead, we tailor our mental models of intelligent agents to involve communication and clarification (versus repeated and/or controlled observations for tools). Natural languages have evolved over time to suit the communication needs of the cultures and context in which they are used, but have historically always been focused on interaction with intelligent agents (namely, humans) [Allen 1995]. While recent UIs have used natural language for interaction, these are often represent changes in modality (e.g., speech), rather than genuine uses of language as one of the most effective ways to mitigate and overcome conceptual and knowledge gaps.

2.2 Collective Intelligence, Human Computation, and Crowdsourcing

Human Computation—which integrates human feedback into computational processes—allows us to leverage input from people in ways never before possible. Crowdsourcing allows these processes to scale up and leverage collective intelligence for more accurate responses [Little et al. 2010; Lin et al. 2012]. Work in this space has shown that these systems can accomplish tasks that computation alone are not able to accomplish, such as image labeling [Russell et al. 2008; Song et al. 2018; Song et al. 2019], creative tasks [Kim et al. 2014; Kittur 2010], and even more general computation [Kittur et al. 2011; Zhang et al. 2011]. However, a majority of these applications focus on tasks in which responses are generated offline for later use (e.g., to train machine learning systems to operate in future interactions).

2.3 Real-time and Interactive Crowdsourcing

Real-time and near-realtime crowdsourcing [Bigham et al. 2010; Lasecki et al. 2011; Bernstein et al. 2011] introduced the ability to elicit human responses on-demand, without needing to collect data a priori in order to train a system. Continuous Real-time Crowdsourcing [Lasecki et al. 2011], or *Interactive Crowdsourcing* [Lasecki et al. 2014], extends this idea further to support continual refinement and progression towards a goal within a single interaction. This has enabled tasks like conversation [Lasecki et al. 2013b; Huang et al. 2016], real-time captioning [Lasecki et al. 2012], robot perception [Gouravajhala et al. 2018], activity and event recognition [Lasecki et al. 2013a; Laput et al. 2015; Lasecki et al. 2018], UI control [Lasecki et al. 2011; Oney et al. 2018] and prototyping [Lasecki et al. 2015; Lee et al. 2017] to be supported on-the-fly with human intelligence. This enables systems to exhibit "*zero-shot functionality*", the ability to work the first time even when a new or unknown situation is encountered.

More recently, Instantaneous Crowdsourcing [Lundgard et al. 2018] has introduced the possibility of human-powered systems capable of responding within milliseconds. This expands the potential applications of collective intelligence to settings where human insight is needed at machine speeds.

3. THE INTELLECTUAL CHALLENGE OF HCOMP FOR INTERACTION

The thesis of this paper is that Hybrid Intelligence Systems (*HyIntS*, pronounced "hints")—which operate via a combination of human (often, crowd) and machine intelligence—have the potential to bridge the socio-technical gap. The core question that must be asked when facilitating interaction using collective intelligence is:

> *How can the varied abilities of multiple intelligent agents be synthesized in a way that is coherent, consistent, and quick enough to enable effective interaction within a target context?*

Scaffolding computational tools with human intelligence can allow for more flexible, and thus robust, systems than we would otherwise be able to automate entirely in the foreseeable future.

3.1 Human-Powered Human-Computer Interaction

As discussed in the Related Work section, prior work in crowdsourcing has demonstrated a wide range of methods and application areas for enabling interaction via human intelligence [Lasecki et al. 2014]. While at first it appears intuitive that humans can collectively facilitate interaction given that individual humans can, the need to *synthesize* responses from multiple people in order to achieve sufficient performance or availability for the system as a whole calls this into question. In fact, many (perhaps most) interactions do not generalize to groups without significant modification [André et al. 2014]. As a result, we need new approaches to coordination that fit within the constraints of interaction.

A central challenge that results from this is the need to find new ways to coordinate the effort of the agents within the system in ways that are able to go beyond traditional limits of human organization. Human Computation has arisen largely as a study of the organization of work and human effort through a computational lens, and has shown promise as a fruitful approach to breaking these prior limits. However, the ways in which people form and update mental models to better communicate, how context is shared effectively between requester and worker, and how communication costs can be mitigated in group interactions are all vital to pushing the envelope in interactive settings.

3.2 Hybrid Intelligence Systems as an Enabling Technology

Combining human and machine intelligence to create Hybrid Intelligence Systems (HyIntS) can leverage human computation's view of organizations through a computational lens to better bridge human effort to computational approaches to solving tasks and coordinating work. Current HyIntS combine human agents with automated tools, but the definition encompasses human-tool and human-agent groups all the same. While a full discussion of the increased complexities of coordinating hybrid groups of agents is beyond the scope of this paper (see [Lasecki 2015] for more on this subject), the challenges largely lay in supporting human reasoning about interaction with non-human agents. This effectively increases the complexity of the interaction surface between contributors within a group.

4. FUTURE DIRECTIONS

Over the last decade, we have seen exceptional strides forward in Human Computation, Crowdsourcing, and other means of leveraging collective intelligence in the facilitation of new capabilities and interactions. However, I believe this is only the beginning – the scale at which we have seen interaction with HyIntS is still relatively small, and history suggests this is but one scale at which to think of interaction. Teams, collaboration, and organization at a macro scale provide a compelling set of future research directions in this space. While much of the recent study of crowdsourcing and collective intelligence in human computation systems has stemmed from the HCI literature in and adjacent-to Computer Science, the study of Management as *technology* (the study of the knowledge and techniques of human organization) is a central precursor to much of this work. Kuipers [Kuipers 2012] notes that in many ways the idea of a corporation—both conceptually and legally—is already one of a single collective entity with capabilities that exceed any one individual. In fact, the value of a firm is defined in terms of the value added to the work of the constituent members, meaning a successful corporation must also outperform the group of employees that comprise it.

Even more recently, work has explored how the idea of collective intelligence at scale can be rethought using a combination of people and technology. The rapid creation of coordinated groups [Retelny et al. 2014] has significant implications for firms [Davis 2018] and organizations broadly. But while interaction at this scale may look different, the core challenge remains the same. Only new work in finding the limits [Retelny et al. 2017] and opportunities can make bridging the interaction gap possible.

5. ACKNOWLEDGEMENTS

I would like to thank Harman Kaur and Stephanie O'Keefe for insightful conversations about these ideas; as well as Sang Won Lee, Sai Gouravajhala, Rebecca Krosnick, Jonathan Kummerfeld, Jean Yong Song, Yan Chen, Lucy Jiang, Jordan Huffaker, Divya Ramesh, John Chung, Alan Lundgard, Jaylin Herskovitz, Ray Fok, Yiwei Yang, Anthony Liu, Mitchell Gordon, Juho Kim, and many others whose work inspired these ideas. This work was generously supported by gratuitous amounts of coffee.

REFERENCES

Mark S. Ackerman. 2000. The intellectual challenge of CSCW: the gap between social requirements and technical feasibility. *Human–Computer Interaction* 15, 2-3 (2000), 179–203.

James Allen. 1995. *Natural language understanding*. Pearson.

Paul André, Robert E. Kraut, and Aniket Kittur. 2014. Effects of simultaneous and sequential work structures on distributed collaborative interdependent tasks. In *ACM Conference on Human Factors in Computing Systems (CHI 2014)*.

Gagan Bansal, Besmira Nushi, Ece Kamar, Dan Weld, Walter S. Lasecki, and Eric Horvitz. 2019. Updates in Human-AI Teams: Understanding and Addressing the Performance/Compatibility Tradeoff. In *AAAI conference on Artificial Intelligence*. AAAI.

Michael S. Bernstein, Joel Brandt, Robert C. Miller, and David R. Karger. 2011. Crowds in Two Seconds: Enabling Realtime Crowd-powered Interfaces. In *ACM Symposium on User Interface Software and Technology (UIST 2011)*.

Jeffrey P. Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C. Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, and Samuel White. 2010. VizWiz: nearly real-time answers to visual questions. In *ACM Symposium on User Interface Software and Technology (UIST 2010)*.

Herbert H. Clark and Susan E. Brennan. 1991. Grounding in communication. *Perspectives on socially shared cognition* 13, 1991 (1991), 127–149.

Gerald F. Davis. 2018. How to communicate large-scale social challenges: The problem of the disappearing American corporation. *Proceedings of the National Academy of Sciences* (2018), 201805867.

James J Gibson. 1977. The theory of affordances. *Hilldale, USA* 1 (1977), 2.

Sai R. Gouravajhala, Jinyeong Yim, Karthik Desingh, Yanda Huang, Odest C. Jenkins, and Walter S. Lasecki. 2018. EURECA: Enhanced Understanding of Real Environments via Crowd Assistance. In *AAAI Conference on Human Computation (HCOMP 2018)*.

Ting-Hao Huang, Walter S. Lasecki, Amos Azaria, and Jeffrey P. Bigham. 2016. "Is there anything else I can help you with?": Challenges in Deploying an On-Demand Crowd-Powered Conversational Agent. In *AAAI Conference on Human Computation*.

Joy Kim, Justin Cheng, and Michael S Bernstein. 2014. Ensemble: exploring complementary strengths of leaders and crowds in creative collaboration. In *ACM Computer supported cooperative work & social computing*. ACM.

Aniket Kittur. 2010. Crowdsourcing, collaboration and creativity. *ACM XRDS: crossroads* 17, 2 (2010), 22–26.

Aniket Kittur, Boris Smus, Susheel Khamkar, and Robert E. Kraut. 2011. CrowdForge: Crowdsourcing Complex Work. In *ACM Symposium on User Interface Software and Technology (UIST 2011)*. 43–52.

Benjamin Kuipers. 2012. An existing, ecologically-successful genus of collectively intelligent artificial creatures. *CI* (2012).

Gierad Laput, Walter S. Lasecki, Jason Wiese, Rovert Xiao, Jeffrey P. Bigham, and Chris Harrison. 2015. Zensors: Adaptive, Rapidly Deployable, Human-Intelligent Sensor Feeds. In *ACM Conference on Human Factors in Computing Systems (CHI 2015)*.

Walter S. Lasecki. 2015. Crowd Agents: Interactive Intelligent Systems Powered by the Crowd. (2015).

Walter S. Lasecki, Mitchell Gordon, Danai Koutra, Malte F. Jung, Steven P. Dow, and Jeffrey P. Bigham. 2018. Glance: Rapidly Coding Behavioral Video with the Crowd. In *ACM Symposium on User Interface Software and Technology (UIST 2014)*.

Walter S. Lasecki, Christopher M. Homan, and Jeffrey P. Bigham. 2014. Architecting Real-Time Crowd-Powered Systems. In *Human Computation Journal (HCJournal, October 2014)*.

Walter S. Lasecki, Juho Kim, Nicholas Rafter, Onker Sen, Jeffrey P. Bigham, and Michael S. Bernstein. 2015. Apparition: Crowdsourced User Interfaces That Come To Life As You Sketch Them. In *ACM Conference on Human Factors in Computing Systems (CHI 2015)*.

Walter S. Lasecki, Christopher Miller, Adam Sadilek, Andrew Abumoussa, Donato Borrello, Raja Kushalnagar, and Jeffrey P. Bigham. 2012. Real-time captioning by groups of non-experts. In *ACM Symposium on User Interface Software and Technology (UIST 2012)*.

Walter S. Lasecki, Kyle I. Murray, Samuel White, Robert C. Miller, and Jeffrey P. Bigham. 2011. Real-time Crowd Control of Existing Interfaces. In *ACM Symposium on User Interface Software and Technology (UIST 2011)*.

Walter S. Lasecki, Young Chol Song, Henry Kautz, and Jeffrey P. Bigham. 2013a. Real-time crowd labeling for deployable activity recognition. In *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, 1203–1212.

Walter S. Lasecki, Rachael Wesley, Jeffrey Nichols, Anand Kulkarni, James F. Allen, and Jeffrey P. Bigham. 2013b. Chorus: A Crowd-Powered Conversational Assistant. In *ACM Symposium on User Interface Software and Technology (UIST 2013)*.

Sang Won Lee, Yujin. Zhang, Isabelle. Wong, Yang Yiwei., Stephanie. O'Keefe, and Walter S. Lasecki. 2017. SketchExpress: Remixing Animations For More Effective Crowd-Powered Prototyping Of Interactive Interfaces. In *ACM Symposium on User Interface Software and Technology (UIST 2018)*.

Christopher Lin, Mausam Mausam, and Daniel Weld. 2012. Dynamically Switching between Synergistic workflows for Crowdsourcing. In *AAAI Conference on Artificial Intelligence*.

Greg Little, Lydia B. Chilton, Max Goldman, and Robert C. Miller. 2010. Turkit: human computation algorithms on mechanical turk. In *ACM Symposium on User Interface Software and Technology (UIST 2010)*. ACM, 57–66.

Alan Lundgard, Yiwei Yang, Maya L. Foster, and Walter S. Lasecki. 2018. Bolt: Instantaneous Crowdsourcing via Just-in-Time Training. In *ACM Conference on Human Factors in Computing Systems (CHI 2018)*.

Donald A. Norman. 2013. *The design of everyday things: Revised and expanded edition*. Basic books.

Donald A. Norman. 2014. Some observations on mental models. In *Mental models*. Psychology Press, 15–22.

Gary M Olson and Judith S Olson. 2000. Distance matters. *Human–computer interaction* 15, 2-3 (2000), 139–178.

Steve Oney, Alan Lundgard, Rebecca Krosnick, Michael Nebeling, and Walter S. Lasecki. 2018. Arboretum and Arbility: Improving Web Accessibility Through a Shared Browsing Architecture. In *ACM Symposium on User Interface Software and Technology (UIST 2018)*. ACM.

Daniela Retelny, Michael S. Bernstein, and Melissa A. Valentine. 2017. No Workflow Can Ever Be Enough: How Crowdsourcing Workflows Constrain Complex Work. *ACM Human-Computer Interaction* 1, CSCW (Dec. 2017).

Daniela Retelny, Sabastien Robaszkiewicz, Alexandra To, Walter S. Lasecki, Jay Patel, Nehar Rahmati, Tulsee Doshi, Melissa Valentine, and Michael S. Bernstein. 2014. Expert Crowdsourcing with Flash Teams. In *ACM Symposium on User Interface Software and Technology (UIST 2014)*.

Bryan C Russell, Antonio Torralba, Kevin P Murphy, and William T Freeman. 2008. LabelMe: a database and web-based tool for image annotation. *International journal of computer vision* 77, 1 (2008), 157–173.

Jean Y. Song, Raymond Fok, Alan Lundgard, Fan Yang, Juho Kim, and Walter S. Lasecki. 2018. Two Tools are Better Than One: Tool Diversity as a Means of Improving Aggregate Crowd Performance. In *ACM Intelligent User Interfaces (IUI 2018)*.

Jean Y. Song, Stephan J. Lemmer, Michael X. Liu, Shiyan Yan, Juho Kim, Jason J. Corso, and Walter S. Lasecki. 2019. Popup: Reconstructing 3D Video Using Particle Filtering to Aggregate Crowd Responses. In *ACM Conference on Intelligent User Interfaces (IUI 2019)*. ACM.

Mark Weiser. 1999. The Computer for the 21st Century. *SIGMOBILE Mob. Comput. Commun. Rev.* 3, 3 (July 1999), 3–11.

JoAnne Yates. 1993. *Control through communication: The rise of system in American management*. Vol. 6. JHU Press.

Haoqi Zhang, Eric Horvitz, Rob C Miller, and David C Parkes. 2011. Crowdsourcing general computation. In *ACM Conference on Human Factors in Computing Systems (CHI 2011)*.