How do Humans Overcome Individual Computational Limitations by Working Together?

Natalia Vélez[a], **Brian Christian**[b], **Mathew Hardy**[c], **Bill D. Thompson**[d], **Thomas L. Griffiths**[c]

[a]Department of Psychology, Harvard University

[b]Center for Human-Compatible AI, University of California, Berekely

[c]Department of Psychology, Department of Computer Science, Princeton University, Princeton, New Jersey, USA

[d]Department of Psychology, University of California, Berekely

Abstract

Since the cognitive revolution, psychologists have developed formal theories of cognition by thinking about the mind as a computer. However, this metaphor is typically applied to individual minds. Humans rarely think alone; compared to other animals, humans are curiously dependent on stores of culturally transmitted skills and knowledge, and we are particularly good at collaborating with others. Rather than picturing the human mind as an isolated computer, we can imagine each mind as a node in a vast distributed system. Viewing human cognition through the lens of distributed systems motivates new questions about how humans share computation, when it makes sense to do so, and how we can build institutions to facilitate collaboration.

Cognitive science has made progress in understanding how humans learn, plan, and act as individuals. However, humans rarely think alone. Compared to other animals, humans are particularly dependent on culturally transmitted skills and knowledge (Henrich, 2015; Sloman & Fernbach, 2018) and motivated to collaborate with others (Tomasello & Hamann, 2012; Tomasello, Melis, Tennie, Wyman, & Herrmann, 2012). Collaboration enables individuals to surmount limitations on their time, cognitive resources, and experience; however, we currently lack a formal framework to describe how humans divide cognitive labor and share the outputs of that labor. This question poses a major challenge to cognitive science.

Since the cognitive revolution, psychologists have developed formal theories of cognition by thinking about the mind as a computer (Gigerenzer & Goldstein, 1996; Newell & Simon, 1972). However, this metaphor is typically applied to individual minds. Constraints

of time, cognitive resources, and experience are *computational* limitations: constraints on the kinds of computational problems individual minds can solve. Perhaps the computer metaphor holds the key to understanding how people work together to take on more complex computational problems. Rather than picturing the human mind as an isolated computer, we can imagine each mind as a node in a vast distributed system. Viewing human cognition through this lens motivates new questions about how humans share computation, when it makes sense to do so, and how we can facilitate collaborations.

Computer scientists have studied distributed systems for decades, analyzing the properties of networks of independent nodes that work together by passing messages (Fokkink, 2018; Lynch, 1996). The Internet is one such system; when you clicked on a link to read this paper, your computer transmitted a request that bounced across computers, allowing you to access content hosted by other computers on the network. In the past, psychologists have used distributed systems as a metaphor for understanding collective social phenomena, such as how collectives develop technological innovations (Smaldino & Richerson, 2013) and store and retrieve memories (Wegner, 1995). Much as how your computer sent a request to another computer to retrieve this paper, you may ask your coworker for the office copier code to "retrieve" that information from their memory.

Moving forward, we propose that there is much to be gained by elaborating the connection between humans and distributed systems. Algorithms for inference or optimization that are designed to be run on distributed systems offer insight into how people might work together to solve similar problems. For example, a sequential Monte Carlo method known as particle filtering suggests one mechanism by which human populations might solve difficult problems of inference through cultural accumulation. In the simplest version, each individual maintains a single hypothesis, with the first generation sampling these hypotheses h from their prior distribution $p(h)$. Each individual in the next generation observes some data d and then randomly chooses individuals from the prior generation until they find one who has a hypothesis consistent with those data, and then adopts that hypothesis. With an infinite population, the probability each individual entertains a specific hypothesis is the posterior probability $p(h|d)$ (Hardy, Krafft, Thompson, & Griffiths, 2022). The population as a whole thus maintains the Bayesian solution, allowing individuals to benefit from accumulated knowledge with little cognitive effort.

One cautionary tale that we can learn from distributed systems is that making the transition from thinking alone to thinking together presents an enormous leap in complexity. Distributed systems can achieve tasks that would be prohibitively expensive for a single computer to do, but they are also more difficult to maintain—individual nodes can fail, messages can get lost, and information can fall out of sync between them. The central challenge of distributed computing is not how to link computers together, but rather how to maintain harmony once they are linked.

Similarly, sharing computation among humans can be enormously beneficial, but it can also introduce new problems. Suppose that Natalia, Brian, Matt, Bill, and Tom decide to collaborate on a paper. By putting their heads together, they can split the work and write a piece that goes beyond any individual author's area of expertise. However, they will also

need to agree on their main points (consensus; Pease, Shostak, & Lamport, 1980), they need to avoid editing the same section at the same time (concurrency, mutual exclusion; Bernstein & Goodman, 1981), they need to deal with situations where one or more authors are not responding (partitions; Gilbert & Lynch, 2002), and they may need to email (asynchronous communication) to find times to join a call (synchronous communication; Charron-Bost, Mattern, & Tel, 1996). Many of these problems have been characterized in machines and are used to design systems that are robust to certain failures. Similarly, taking inspiration from distributed systems may provide a principled way to predict when it makes sense to think together—namely, when the gain in computational power outweighs the costs of the problems it causes.

This example also illustrates that human-distributed computation is often enabled by machines and cultural institutions. Without telecommunications, computer networks, online word processors, and email services, the authors would not be able to edit the same document simultaneously. It is therefore important to understand how cultural technologies and institutions affect the dynamics of human-distributed systems (Gabora, 2013; Hutchins, 1995), and how to best design services to enable people to work together (Maglio & Spohrer, 2008). The last decade has revealed some of the unintended consequences of large-scale social networks and recommendation systems, including increased polarization and the creation of "filter bubbles" (Pariser, 2011). Participating in a social network can allow people to make decisions informed by social information but can also magnify biases when that information comes from people who are themselves biased. By viewing the social network as a system that executes an algorithm based on human decisions, we have the opportunity to ask how that algorithm can be improved. Making a small tweak to the algorithm—making it more likely to share decisions that are more representative of an unbiased population—removes bias magnification while preserving the benefits of participating in a social network (Hardy, Thompson, Krafft, & Griffiths, 2022).

Taking inspiration from distributed systems of machines can be a rich source of hypotheses about how humans share computation; however, ultimately, we need theories of distributed computation that are tailored to the capabilities and constraints of human minds. When extended to the natural world, the challenges that are characteristic of distributed systems are compounded by the fact that the "messages" that biological intelligences pass to one another are constructed in social contexts. Unlike machines, humans do not share the contents of their minds directly, and we do not communicate using protocols that are imposed by a designer. Instead, humans rely on language, gesture, and other means to share their thoughts. We can quickly develop linguistic conventions to communicate more efficiently about a task (Fusaroli et al., 2012; McCarthy, Hawkins, Wang, Holdaway, & Fan, 2021) and rely on our ability to reason about other minds to unpack the meaning behind other people's messages (Vélez & Gweon, 2019, 2021). Further, humans can organize in various ways to achieve collaborative goals, creating labs, unions, and hunting parties. Understanding how humans form social structures in response to particular problems presents a daunting challenge (Ostrom, 2009); groups may exhibit emergent characteristics —such as spontaneous subdivisions or their own forms of collective intentionality—that were not designed by any one individual (Goldstone & Janssen, 2005; Tollefsen, 2006). These aspects of communication and social organization create distinctive forms of shared

computation. This, then, is our puzzle: understanding the unique properties of human-distributed systems.

References

Bernstein PA, & Goodman N (1981). Concurrency control in distributed database systems. ACM Computing Surveys, 13(2), 185–221.

Charron-Bost B, Mattern F, & Tel G (1996). Synchronous, asynchronous, and causally ordered communication. Distributed Computing, 9(4), 173–191.

Fokkink W. (2018). Distributed algorithms, second edition: An intuitive approach. Cambridge, MA: MIT Press.

Fusaroli R, Bahrami B, Olsen K, Roepstorff A, Rees G, Frith C, & Tylén K (2012). Coming to terms: Quantifying the benefits of linguistic coordination. Psychological Science, 23(8), 931–939. [PubMed: 22810169]

Gabora L. (2013). Cultural evolution as distributed computation. Available at: http://arxiv.org/abs/1310.6342. Accessed October 31, 2022.

Gigerenzer G, & Goldstein DG (1996). Mind as computer: Birth of a metaphor. Creativity Research Journal, 9(2-3), 131–144.

Gilbert S, & Lynch N (2002). Brewer's conjecture and the feasibility of consistent, available, partition-tolerant web services. SIGACT News, 33(2), 51–59.

Goldstone RL, & Janssen MA (2005). Computational models of collective behavior. Trends in Cognitive Sciences, 9(9), 424–430. [PubMed: 16085450]

Hardy MD, Krafft PM, Thompson B, & Griffiths TL (2022). Overcoming individual limitations through distributed computation: Rational information accumulation in multigenerational populations. Topics in Cognitive Science, 14(3), 550–573. [PubMed: 35032363]

Hardy MD, Thompson BD, Krafft PM, & Griffiths TL (2022). Bias amplification in experimental social networks is reduced by resampling. Available at: http://arxiv.org/abs/2208.07261. Accessed October 31, 2022.

Henrich J. (2015). The secret of our success. Princeton, NJ: Princeton University Press.

Hutchins E (1995). Cognition in the wild. Cambridge, MA: MIT Press.

Lynch NA (1996). Distributed algorithms. Burlington, MA: Morgan Kaufmann Publishers.

Maglio PP, & Spohrer J (2008). Fundamentals of service science. Journal of the Academy of Marketing Science, 36(1), 18–20.

McCarthy WP, Hawkins RD, Wang H, Holdaway C, & Fan JE (2021). Learning to communicate about shared procedural abstractions. Proceedings of the 43rd Annual Meeting of the Cognitive Science Society. Vienna, Austria (pp. 77–83). https://rxdhawkins.files.wordpress.com/2021/05/caml_cogsci-1.pdf

Newell A, & Simon HA (1972). Human problem solving. Englewood Cliffs, NJ: Prentice-Hall.

Ostrom E. (2009). Understanding institutional diversity. Englewood Cliffs, NJ: Princeton University Press.

Pariser E. (2011). The filter bubble: What the Internet is hiding from you. London, UK: Penguin Books.

Pease M, Shostak R, & Lamport L (1980). Reaching agreement in the presence of faults. Journal of the ACM, 27(2), 228–234.

Sloman S, & Fernbach P (2018). The Knowledge illusion: Why we never think alone. New York, NY: Riverhead Books.

Smaldino PE, & Richerson PJ (2013). Human cumulative cultural evolution as a form of distributed computation. In Michelucci P (Ed.), Handbook of human computation (pp. 979–992). New York: Springer.

Tollefsen DP (2006). From extended mind to collective mind. Cognitive Systems Research, 7(2), 140–150.

Tomasello M, & Hamann K (2012). Collaboration in young children. Quarterly Journal of Experimental Psychology (2006), 65(1), 1–12. [PubMed: 22171893]

Tomasello M, Melis AP, Tennie C, Wyman E, & Herrmann E (2012). The interdependence hypothesis. Current Anthropology, 53(6), 673–692.

Vélez N, & Gweon H (2019). Integrating incomplete information with imperfect advice. Topics in Cognitive Science, 11(2), 299–315. [PubMed: 30414253]

Vélez N, & Gweon H (2021). Learning from other minds: An optimistic critique of reinforcement learning models of social learning. Current Opinion in Behavioral Sciences, 38, 110–115. [PubMed: 35321420]

Wegner DM (1995). A computer network model of human transactive memory. Social Cognition; New York, 13(3), 319–339.