Cognitive Science of Augmented Intelligence

Marina Dubova[1], Mirta Galesic[2,3,4], and Robert L. Goldstone[1,5]

[1]Cognitive Science Program, Indiana University, Bloomington, IN 47405, USA

[2]Santa Fe Institute, Santa Fe, NM 87501, USA

[3]Complexity Science Hub Vienna, 1080 Vienna, Austria

[4]Vermont Complex Systems Center, University of Vermont, Burlington, VT 05405, USA

[5]Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405, USA

mdubova@iu.edu

Declarations of interest: none.

Correspondence concerning this article should be addressed to Marina Dubova, Cognitive Science Program, Indiana University, 1101 E. 10th Street, Bloomington, IN 47405 USA

Keywords: cognitive science, augmented intelligence, extended cognition, collective cognition, social learning, human-machine collaboration

Abstract

Cognitive science has been traditionally organized around the individual as the basic unit of cognition. Despite developments in areas such as communication, human-machine interaction, group behavior, and community organization, individual-centric approach heavily dominates both cognitive research and its application. A promising direction for cognitive science is the study of augmented intelligence, or the way social and technological systems interact with and extend individual cognition. The cognitive science of augmented intelligence holds promise in helping society tackle major real-world challenges that can only be discovered and solved by teams made of individuals and machines with complementary skills who can productively collaborate with each other.

1. Minds as augmented by other minds and technologies

Cognitive science is changing fast, but is still anchored in its tradition of studying how people learn and cognize on their own. Most of cognitive science studies mechanisms of individual perception, learning, and decision making, rarely considering humans interacting with each other and with the technologies they devise. To strengthen a recent rise of proposals to move away from the individual-centric cognitive science (e.g. Sloman & Fernbach, 2018; Pickering & Garrod, 2021; Chater & Loewenstein, 2022; Rajaram, 2022), we argue for expanding the subject of cognitive science beyond the individual to cognitive systems of minds and technologies.

We believe that a major direction of progress for cognitive science will be augmented intelligence – understanding how people recruit other people and technologies to improve their thinking, and how these same people, in turn, are recruited into larger cognitive systems. Humans rarely solve problems on their own (Clark, 1998; 2008; Clark & Chalmers, 1998). Rather, we learn from and in coordination with other people, and we create tools to substitute and extend our cognitive processing. For instance, we often rely on others' knowledge to solve our problems (Sloman & Fernbach, 2018) and self-organize in teams to make decisions about complicated issues. And, novel digital technologies from search engines to crowdsourcing platforms and AI systems assist in our individual and collective learning and decision making. These are not the only ways to conceptualize augmented intelligence. In fact, when many people consider augmented intelligence, they imagine neurotropic pharmaceuticals or brain implants. These are exciting, distant possibilities, but focusing on these risks ignoring the immediately available ways in which our minds are already being extended by the social and technical networks we inhabit.

2. Cognitive Science of Augmented Intelligence: Why

Expanding cognitive science to include systems larger than single individuals is not just a methodological exercise. It will facilitate our understanding of individual, social, and human-machine behavior. Cognitive scientists have already made progress in understanding one-person cognitive systems. We have designed methods, constructed theories, and built models to gain insights into individual cognition. Applying these developments to multi-agent and cyborg intelligent systems can jumpstart the science of intelligent systems in the broader sense. Reciprocally, larger, possibly intelligent systems, such as social media communities, leave behind a readily inspectable digital trail of artifacts, so they can eventually inform our understanding of our own minds. We can often study the development history and the interactions of actors within such systems with higher precision than we are able to study human cognitive development over the lifespan or interactions of neurons in our brains. Moreover, intelligent groups often involve dyads, tens, or hundreds of actors, which enables the development of models and tools in a manageable context that is less challenged by scale than studying the human brain involving interactions of billions of cells (Navas-Zuloaga, Pavlic & Smith, 2022).

Studying cognition extended beyond an individual human could fulfill a long-standing ambition of cognitive science – to understand the principles of intelligent systems in their generality, not tied to specific incarnations such as individual organisms defined by their physical boundaries (Clark, 2008). Human beings are undoubtedly fine specimens of cognitive systems. By extending mechanisms shown to be at work in human cognition to groups and human-machine teams, we can explore the necessary ingredients and scope of these mechanisms. To understand what is shared by all languages, it is better to study Mandarin, Swahili, and Dutch than French, Italian, and Spanish. Likewise, if we are to develop a general science of intelligent systems, it behooves us to study it in its diverse manifestations at different biological, social, and technological scales. By triangulating on cognitive systems from these multiple vantage points, we can come to better understand their essential nature.

3. Cognitive Science of Augmented Intelligence: How

Recognizing cognitive science as a science of intelligent systems broadly construed does not mean that anything counts as cognitive. Instead, the augmented intelligence perspective invites the field to articulate new criteria for cognition beyond the self-centered and shallow criterion of being a *homo sapiens* (Huebner, 2014). Reengineering the notion of "cognitive system" can proceed in a bottom-up or top-down way. A top-down approach involves systematizing what we already know about individual cognition and identifying its most fundamental properties. Possible top-down criteria include adaptation as a result of competition, selection, coordination (e.g. Galesic et al., 2022), the emergence of specialized modules for information processing (Goldstone, 2019; Goldstone & Theiner, 2017), non- or near-decomposability of a system's functionality to smaller units (Simon, 1962; e.g. cognitive function is not reducible to one brain area; collective action is not reducible to individual actions), or the specific mechanisms (e.g. memory, perception) or contents that ostensibly constitute individual cognition (e.g. Sloman, Patterson & Barbey, 2021). The bottom-up approach to reengineering the "cognitive system" starts by postulating certain systems as cognitive (e.g. individual humans, teams and corporations, human-machine collaborations, AI systems), and studying the properties that they share.

Candidate cognitive systems have been usefully employed to understand systems wider than, narrower than, and simply different from single central nervous systems. For example, in competitive specialization, a group of originally homogeneous, undifferentiated units develops towards a self-organized division of labor among the units (Rumelhart & Zipser 1985). In diffusion-based decision-making, information accumulation until a certain threshold is reached effectively captures the details of individual neurons' firing rates (Gold & Shadlen, 2007) and behavior in individual human cognitive tasks such as recognition memory and categorization (Bogacz et al. 2006; Forstmann, Ratcliff, & Wagenmakers, 2016), but also collective phenomena (Marshall et al., 2011). In temporal difference learning, learning is based not only on a rare external reward but also on the degree to which the system can predict that reward. The prediction error assessment can be supplied by the neurotransmitter dopamine in nervous systems (Schultz, Dayan, & Montague, 1997), by people tasked with reviewing the performance of others in team contexts, or by machine-implemented reinforcement algorithms (Sutton & Barto, 1998). In similarity-based sampling, apt behaviors for a given situation are decided on by sampling previously remembered similar situations. This notion has been used to model individual financial decisions (Stewart, Chater, & Brown 2006), categorization judgments (Nosofsky, 1984), and social judgments (Galesic, Olsson, & Rieskamp, 2018; Pachur, Hertwig, & Rieskamp, 2013), as well as to develop inductive algorithms such as Support Vector Machines and Bayesian inference. In a collective context, situations could be sampled not only from an individual's own memory, but also from others (Hirst, Yamashiro, & Coman, 2018). Conversations, stories, movies, and gossip all fulfill the function of helping people to behave better by learning from the real or imagined experiences of others.

Some recent developments demonstrate an ongoing transition from an individual-centric to augmented intelligence perspective for tackling real world issues. For example, Artificial Intelligence (AI) research has been increasingly adopting a human-machine coordination perspective instead of focusing on either human or machine learning individually. The eXplainable AI initiative (Gunning et al., 2019) is one example of such an approach. This field focuses on developing algorithms that are understandable by the human user (Rudin, 2019) or interpreting the "reasoning" of black box AI systems so that humans can use their advice more effectively (Gunning et al., 2021). Human-AI complementarity explores a broad range of conditions that contribute to effective teaming of artificial and human minds. Here, the unique strengths of human and artificial intelligences allow them to combine in more successful teams than would be achieved by either human-human or AI-AI teams (Steyvers et al., 2021). Other work explores ways to use AI to augment human cognition in particular high stakes domains, such as applying machine denoising methods to notice and correct inconsistencies in medical decisions (Hasan et al., 2022).

4. Outlook

Studying intelligence as augmented by social connections and technologies is particularly timely for several reasons. First, there have been important developments in the formal modeling of collective behavior and human-machine interaction. Second, recent technological advances enable scientists to explore human-human and human-technology interaction with unprecedented precision both in a laboratory and by harvesting real-world data. Perhaps most importantly, humanity is becoming increasingly more socially and technologically interconnected. Mass-produced scholarly work such as Wikipedia demonstrates that a large and decentralized collective can produce highly structured and high quality information. Social media sites such as Facebook and Twitter have both positive and negative individual and societal effects that can be difficult to anticipate (Lewandowsky et al., 2020). Machines show great promise for partnering with people for problem solving in such areas as crowdsourcing, virtual worlds, and AI-supported educational design. The rise of adaptive technologies that mediate human social interactions opens up new modes of sociality to be investigated. Therefore, a promising future of cognitive science involves studying how individual cognition gives rise to and is in turn shaped by collectives from families to political parties and countries, and technologies from search algorithms to virtual assistants and self-driving cars. To better design or, at least, monitor these systems as they evolve, we need better understanding of cognitive mechanisms that underlie them.

Acknowledgement
This research has been supported by the National Science Foundation, Program in the Science of Learning and Augmented Intelligence Award #193683.

1. Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced choice tasks. Psychological Review, 113, 700–765.
2. Chater, N., & Loewenstein, G. (2022). The i-frame and the s-frame: How focusing on the individual-level solutions has led behavioral public policy astray. *Available at* SSRN 4046264.
3. Clark, A. (1998). *Being there: Putting brain, body, and world together again*. MIT press.
4. Clark, A. (2008). Supersizing the mind: Embodiment, action, and cognitive extension. Oxford University Press.
5. Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7–19.
6. Forstmann, B. U., Ratcliff, R., & Wagenmakers, E. J. (2016). Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. Annual Review of Psychology, 67, 641–666.
7. Galesic, M., Barkoczi, D., Berdahl, A., Biro, D., Carbone, G., Giannoccaro, I., ... & Stein D. L. (2022). *Beyond collective intelligence: Collective adaptation.*
8. Galesic, M., Olsson, H., & Rieskamp, J. (2018). A sampling model of social judgment. *Psychological Review*, *125*(3), 363
9. Gold, I. J., & Shadlen, M. N. (2007). The neural basis of decision making. Annual Review of Neuroscience, 30, 535–574.
10. Goldstone, R. L. (2019). Building Inner Tools. *Andy Clark and His Critics*, 113.
11. Goldstone, R. L., & Theiner, G. (2017). The multiple, interacting levels of cognitive systems (MILCS) perspective on group cognition. *Philosophical Psychology*, 30(3), 338–372.
12. Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G. Z. (2019). XAI—Explainable artificial intelligence. *Science robotics*, 4(37), eaay7120.
13. Gunning, D., Vorm, E., Wang, J. Y., & Turek, M. (2021). DARPA's explainable AI (XAI) program: A retrospective. *Applied AI Letters*, 2(4), e61.
14. Hasan, E., Eichbaum, Q., Seegmiller, A. C., Stratton, C., & Trueblood, J. S. (2022). Improving Medical Image Decision-Making by Leveraging Metacognitive Processes and Representational Similarity. *Topics in Cognitive Science*, 14(2), 400-413.
15. Hirst, W., Yamashiro, J. K., & Coman, A. (2018). Collective memory from a psychological perspective. *Trends in cognitive sciences*, 22(5), 438-451.
16. Huebner, D. (2014). Macrocognition: A theory of distributed minds and collective intentionality. Oxford University Press. Chapters 1-2.
17. Lewandowsky, S., Smillie, L., Garcia, D., Hertwig, R., Weatherall, J., Egidy, S., ... & Leiser, M. (2020). Technology and democracy: Understanding the influence of online technologies on political behaviour and decision-making.
18. Marshall, J. A. R., Bogacz, R., Dornhaus, A., Plaque, R., Kovacs, T., & Franks, N. R. (2011). On optimal decision-making in brains and social insect colonies. Journal of the Royal Society Interface, 6, 1065–1074.

19. Navas-Zuloaga, M. G., Pavlic, T. P., & Smith, B. H. (2022). Alternative model systems for cognitive variation: eusocial-insect colonies. *Trends in Cognitive Sciences*.
20. Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, memory, and cognition*, 10(1), 104.
21. Pachur, T., Hertwig, R., & Rieskamp, J. (2013). The mind as an intuitive pollster: Frugal search in social spaces. In *Simple heuristics in a social world* (pp. 261-291). Oxford University Press.
22. Pickering, M. J., & Garrod, S. (2021). *Understanding dialogue: Language use and social interaction*. Cambridge University Press.
23. Rajaram, S. (2022). Collective memory and the individual mind. *Trends in Cognitive Sciences*.
24. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206-215.
25. Rumelhart, D. E., & Zipser, D. (1985). Feature discovery by competitive learning. *Cognitive science*, 9(1), 75-112.
26. Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275(5306), 1593-1599.
27. Simon, H. A. (1962). The Architecture of Complexity. *Proceedings of the American Philosophical Society*, 106(6), 467–482.
28. Sloman, S., & Fernbach, P. (2018). *The knowledge illusion: Why we never think alone*. Penguin.
29. Sloman, S. A., Patterson, R., & Barbey, A. K. (2021). Cognitive neuroscience meets the community of knowledge. *Frontiers in Systems Neuroscience*, 120.
30. Stewart, N., Chater, N., & Brown, G. D. (2006). Decision by sampling. *Cognitive psychology*, 53(1), 1-26.
31. Steyvers, M., Tejeda, H., Kerrigan, G., & Smyth, P. (2022). Bayesian modeling of human–AI complementarity. *Proceedings of the National Academy of Sciences*, 119(11), e2111547119.
32. Sutton, R. S., & Barto, A. G. (1998). Introduction to reinforcement learning.