How can Machines be "Team Members"?

R.R. Hoffman
March 2020

rhoffman@ihmc.us

A Little Story

While on a recent driving trip, I beheld a field of short bushes with reddish-brown leaves. Unsure of their identity, I Googled to ask about the crops grown in that region. After receiving a response, I said back to Google, "Thank, you Google." Google replied, "You are welcome, we aim to please." Why did I laugh? Obviously, there was no genuine agency here, no intention, no emotion. Merely a search to find word matches, which mapped to a look-up table that pointed to an audio file.

While I will take a measure of responsibility for injecting the notion of "team players" into the modern discourse, I now plead guilty to brain fog, and plead for a clearer understanding of this notion. The concept that computers or AI systems can be "team members" has engulfed many research programs. This is not good, and in this essay I try to express why.

Seminal work on this topic (e.g., Johnson, et al., 2014; Klein, et al., 2004) proposed a set of requirements or constraints on what it would mean for a computer to be a team *player*. This work adduced some crucial and powerful ideas, including the notion that the machine ("agent") must be observable, directable, and predictable. "Intelligent agents must be able to adequately model the other participants' intentions and actions" (Klein, et al., 2004, p. 92). These are constraints on the humans and on the machines, hence the generic term "agent."

In this essay I do not wish to detract from these principles in any way. I note that the Klein, et al. essay carefully referred to machines as team *players*, not team*mates* or team *members*. This distinction is crucial. Words matter.

Let's start with "agent." This computer science term clearly has the implication of giving the machine agency, which is a philosophical leap. The "seven cardinal virtues" of human-machine teamwork (Johnson, et al., 2014; Klein, et al., 2004) could be recast, substituting the word "machine" or the word "computer" for the word "agent" (or "robot," or the phrase "team member") and absolutely nothing would be lost of the core content or intent of these important principles.

The notion of machines as team *players* has been taken to mean that the machine will be a team *member*. The machine will be a human-like thinker and actor. Such is the effect of viral catch-phrases, which engenders mythos and is ultimately misleading. And ultimately dangerous. When people assume that their machine "teammates" are really human-ish, and then the machine makes a sort of error that a human would never make then something bad happens, including the rapid evaporation of trust in the machine.

Not long ago there were a number of programs aimed at creating intelligent "associates," such as the Pilot's Associate (Banks and Lizza, 1991) and the Intelligence Analyst's Associate (Chappell, et al., 2004; Fikes, Ferrucci, & Thurman, 2005). The shift from referring to machine associates to machine teammates has insinuated current discourse on AI and technology, including numerous government funding programs, clarion calls for "cyber teammates" (e.g., Buyonneau and Le Dez, 2019), and a considerable amount of research on how college students interact with computers (e.g., Nass, et al., 1996). An effort is underway to establish a "Human-Robot Teaming" Technical Group within the Human Factors and Ergonomics Society.

This work is all potentially valuable. However, it gets tarnished by the recent hype claiming that more computing power and deep learning will solve all our problems. And especially the hype that AI is an extension of the brain.

Across the pertinent conference proceedings and journals there are scores upon scores of articles that use such phrases as "human-machine teams," "robotic teammates," "human-machine collaboration," and even "cognitive cooperation." Rarely does one find a paper that wonders about the appropriateness of conceiving of a computer or robot as a human-ish teammate. Two papers that do are one by William Clancey (2004) in which he questions the anthropomorphism and boldly asserts that in order for a computational system to really be a genuine teammate it will require consciousness. Similarly, Groom and Nass (2007) question what it means to think of robots as teammates, arguing that since they lack humanlike mental models and a sense of self, robots will be untrustworthy and will fail to satisfy the requirements for a humans to be teammates.

> Robots would not be used for side-by-side interaction if [people] did not believe that robots have assets to contribute that humans do not. The human tendency to see "humanness" everywhere has led researchers to impose a model of interaction suitable only for human-human interaction (p. 496).

To the point of this blog, I have never heard any statement made at a meeting, or any statement included in a funded program description, where someone calls out the anthropomorphism, let alone expresses any caution about it. It seems to be just taken for granted that the machine will be a human-ish teammate, and achieving this means making the machine agentive.

Metaphors serve a number of roles in scientific reasoning (Hoffman, 1980). This includes suggesting testable hypotheses and experimental designs. But metaphors are not substitutes for fully-formed scientific theories; they are always incomplete, incorrect, or misleading in some respects Indeed, this is a virtue of the metaphors: Apperception of the ways in which they are incorrect leads to scientific advances. I assert that the notion of a computer or robot as a "team member" is a misleading metaphor; dangerous in that it tacitly attributes capabilities and qualities that the machine simply does not have, and likely will not have for some time to come.

Machines are tools. They cannot be team members in any genuine human sense. Here's the key point: Machines can be (and are being) made to act ***as if*** they are team players. But this will be only in certain respects, only in certain stable contexts, and only with respect to certain tasks or subtasks. The machine's teaminess will be brittle and transient because the world is not fixed. And

the machine only has agency to the extent that its design is a stand-in for the intentions of its human designers.

We all want better machines. But we also need to be mindful of how our terminology, mis-attributions, anthropomorphisms, and mindless metaphors can mislead, however serviceable they may be as hyperbolic clarion calls.

References

Banks, S.S., & Lizza, C.S. (1991). Pilot's Associate: A cooperative, knowledge-based system application. *IEEE Expert, 6*, 18-29.

Chappell, A.R., Cowell, A.J., Thurman, D.A., & Thomas, J.R. (2004). Supporting mutual understanding in a visual dialog between analyst and computer. In *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meetin*g (pp. 376-380). Santa Monica, CA: Human Factors and Ergonomics Society.

Clancey, W. J. (2004) Roles for agent assistants in field science: Understanding personal projects and collaboration. *IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews, 34* (2) Special Issue on Human-Robot Interaction, pp. 125-137.

Fikes, R., Ferrucci, D., & Thurman, D. (2005). Knowledge associates fort novel intelligence. Presentation at the International Conference on Intelligence Analysis. Washington, DC: Office of the Assistant Director of National Intelligence.

Groom, V. and Nass, C. (2007). Can robots be teammates? *Interaction Studies, 8*, 483-500.

Hoffman, R.R. (1980). Metaphor in science. In R. P. Honeck and R. R. Hoffman (Eds.), *Cognition and figurative language* (pp. 393-423). Mahwah, NJ: Erlbaum.

Klein, G., Woods, D.D., Bradshaw, J.D., Hoffman, R.R. and Feltovich, P.J. (November/December 2004). Ten challenges for making automation a "team player" in joint human-agent activity. *IEEE Intelligent Systems*, pp. 91-95.

Johnson, M., Bradshaw, J.M.,. Feltovich, P.J.,. Hoffman, R.R., Jonker, C., & van Riemsdijk, B. (May/June 2011). Beyond Cooperative Robotics: The Central Role of Interdependence in Coactive Design. *IEEE Intelligent Systems*, pp. 81-88.

Johnson, M., Bradshaw, J.M., Hoffman, R.R., Feltovich, P.J. & Woods, D.D. (November/December 2014). Seven cardinal virtues of human-machine teamwork. *IEEE Intelligent Systems*, pp. 74-79.

McBride, N., and Hoffman, R.R. (2016, September/October). Bridging the ethical gap: From human principles to robot instructions. *IEEE Intelligent Systems*, pp. 76-82.

Nass, C., Fogg, B.J., & Moon, Y. (1996). Can computers be teammates? International Journal of Human-Computer Studies, 45, 669-678.