KEY CRITERIA AND CHALLENGES FOR HUMAN-AI/AUTONOMY TEAMING

Nathan J. McNeese
College of Engineering, Computing and Applied Sciences Dean's Professor
Assistant Professor of Human-Centered Computing
Clemson University

College of ENGINEERING, COMPUTING AND APPLIED SCIENCES

TRACE RESEARCH GROUP

Human-AI Teaming: In General

TRACE RESEARCH GROUP

Human–Autonomy Teaming: A Review and Analysis of the Empirical Literature

Thomas A. O'Neill, University of Calgary, Canada, Curtin University, WA, Australia, **Nathan J. McNeese**, Clemson University, South Carolina, USA, **Amy Barron**, University of Waterloo, Canada, and **Beau Schelble**, Clemson University, South Carolina, USA

Objective: We define human–autonomy teaming and offer a synthesis of the existing empirical research on the topic. Specifically, we identify the research environments, dependent variables, themes representing the key findings, and critical future research directions.

Background: Whereas a burgeoning literature on high-performance teamwork identifies the factors critical to success, much less is known about how human–autonomy teams (HATs) achieve success. Human–autonomy teamwork involves humans working interdependently toward a common goal along with autonomous agents. Autonomous agents involve a degree of self-government and self-directed behavior (agency), and autonomous agents take on a unique role or set of tasks and work interdependently with human team members to achieve a shared objective.

Method: We searched the literature on human–autonomy teaming. To meet our criteria for inclusion, the paper needed to involve empirical research and meet our definition of human–autonomy teaming. We found 76 articles that met our criteria for inclusion.

Results: We report on research environments and we find that the key independent variables involve autonomous agent characteristics, team composition, task characteristics, human individual differences, training, and communication. We identify themes for each of these and discuss the future research needs.

Conclusion: There are areas where research findings are clear and consistent, but there are many opportunities for future research. Particularly important will be research that identifies mechanisms linking team input to team output variables.

Keywords: teamwork, team processes, team performance, human–autonomy teaming, human–automation interaction, human–agent collaboration

Address correspondence to Thomas A. O'Neill, Department of Psychology, University of Calgary, AB, T2N 1N4, Canada; e-mail: toneill@ucalgary.ca

HUMAN FACTORS
Vol. 00, No. 01, Month XXXX, pp. 1–35
DOI:10.1177/0018720820960865
Article reuse guidelines: sagepub.com/journals-permissions

INTRODUCTION

Recently, research has been using the term human–autonomy teams (HATs) to describe humans and intelligent, autonomous agents working interdependently toward a common goal (Chen et al., 2016; Johnson et al., 2012; Wynne & Lyons, 2018). HAT has been described as at least one human working cooperatively with at least one autonomous agent (McNeese et al., 2018), where an autonomous agent is a computer entity with a partial or high degree of self-governance with respect to decision-making, adaptation, and communication (Demir et al., 2016; Mercado et al., 2016; Myers et al., 2019). As noted by Larson and DeChurch (2020, p. 10), "we are quickly approaching a time when digital technologies are as agentic as are human counterparts." With continuous advancements in artificial intelligence (AI), autonomous agents can perform a greater number of dynamical functions in both teamwork and taskwork than ever before (Seeber et al., 2020), and they are beginning to be viewed as teammates rather than tools (Grimm et al., 2018a; Lyons et al., 2018). For example, autonomous agents can increasingly participate in teamwork activity involving coordination, task reallocation, and continuous interaction with humans and other autonomous agents (Chen et al., 2016; Johnson et al., 2012; Shannon et al., 2017).

The potential of autonomous agents working with humans opens up an interesting question, which involves both articulating a clear definition of HATs as well as identifying the factors that make these teams successful. Indeed, empirical research on HATs is burgeoning. Yet, the findings remain scattered, thereby obscuring a clear perspective on the state of the science. Accordingly, an integrative review is needed, which we offer here. By conducting a review

Open Access!!!

Definition: Autonomous technology (independent of human control) collaborating with humans as teammates to undertake taskwork and teamwork functions (McNeese et al., 2018; O'Neill, McNeese, Barron, Schelble, 2022).

76 empirical studies meeting HAT criteria:

- Team is 2 or more members working **interdependently** toward common goal
- At least one autonomous agent…
- …where the autonomous agent(s) occupy unique roles on the team and possess a significant **degree of agency.**

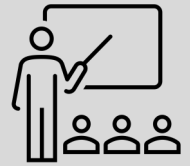
Education and Training in Human-AI Teaming

Let's Not Forget About Education….

- Problems with technical/AI literacy
- Humans see AI teammates as fundamentally different from human teammates (Zhang, McNeese et al., 2021).
- Understanding of human-AI teaming both conceptually and practically is a critical need in research moving forward.

Addressing These Issues Then Informs Training Development

- *Types of Training*
    - Cross training (Volpe et al., 1996; Nikolaidis & Shah, 2013), procedural (task) training (Gorman et al., 2010), adaptive/perturbation training (Ramakrishnan et al., 2017).
- *Training Content*
    - Simulation based training is highly effective given the importance of context (Marlow et al., 2017).
    - Taskwork and teamwork serve as the most relevant content teams train on, however, education regarding AI literacy is necessary.
- *Research Needs*
    - (We Need More Training Research!) Developing human-AI specific training content and calibrating human expectations of AI teammates.



Trust in Human-AI Teaming

Defining Trust (Its Complicated)

- *Between Humans*
  - “The willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor” (Mayer et al., 1995).
- *Between Humans and Systems*
  - “The attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty” (Lee & See, 2004).

Trust Related Findings

- Lower trust in an AI teammate is associated with reduced team performance and reduced trust in human teammates (McNeese et al., 2021).
- Unethical actions by an AI teammate reduce trust in the AI (Schelble, Lopez, Textor, Zhang, McNeese et al., 2022).
- Becoming the only human member of a human-AI team reduced trust in the AI teammate (Schelble, Flathmann, McNeese et al., 2022).

More work to be done in this space:

- Multi-dimensional data collection (subjective and objective) of teaming inputs of trust, breaking trust down into more manageable sub-concepts (not just the general concept of ”trust”), and real-time measurements

Performance in Human-AI Teaming

Operationalizing Human-AI Team Performance

- *Objective Measures*
  - Composite Team Score: Time, Accuracy, Objectives Completed, Accuracy, Resource Usage (McNeese et al., 2018)
  - Individual Score (Schelble, Flathmann, McNeese et al., 2022)
- *Subjective Measures*
  - Perceived Team Effectiveness Survey (Rentsch et al., 2002)
  - Workload (Hart & Staveland, 1988)

Performance Related Findings

- Depends on team and context, much like human-human teams.
- Higher levels of agent autonomy performance tends to result in better outcomes.
- Mixed human-AI teams outperformed human-only teams (McNeese et al., 2021), though this is highly related to task complexity (McNeese et al., 2018).
- Generally, HATs tend to lag in performance relative to teams comprised only of humans. Mainly due to AI agents not being designed with teaming behaviors.

As a research domain, Human-AI Teaming is incredibly healthy and continually growing fast.

More questions than answers, at this stage:

-Conceptual/Definitional

-Empirical

-Design

-Implementation

But…

Human-AI teaming can spiral out of control without scoping and realistic expectations.

Human-AI teaming is not and was never meant to be viewed in the same exact light as human-human teaming.

Sets limits and unrealistic expectations on what H-AI teaming can and should be. H-AI teaming requires an open mind to new teaming behaviors and being open to differences from H-H teaming. We need to embrace these differences.

An AI teammate does not imply humanness.

An AI teammate will not and should not interact in the same way human teammates do. That doesn't automatically make it a tool.

Potential Next Steps:

Continue to explore if models of human-human teaming and human-animal teaming are effective and/or appropriate for HAT.

Focus on continuing to define/computationally operationalize what an AI teammate can and should do in the near and long term.

Explore additional means of communication in addition to natural language processing.

Explore the potential of a theory of behavior rather than a theory of mind for AI teammates.

Develop HAT training materials and environments for the purpose of allowing humans to experience realistic HAT and calibrate their expectations accordingly.

H-A teaming is not just a metaphor to explain human-AI interaction. AI are already interacting with humans in collaborative environments where teaming behavior is expected.

We need some (not all) AI to broadly understand teaming behavior right now, not a decade from now.

mcneese@clemson.edu