Adopting the Theory of Distributed Cognition for Human-AI Cooperation

MATTHEW SIDJI, University of Melbourne, Australia
WALLY SMITH, University of Melbourne, Australia
MELISSA J. ROGERSON, University of Melbourne, Melbourne

Questions around human-AI cooperation are increasingly important as AI begins to act as a teammate as opposed to as a tool used by human operators. To effectively design for this, a deeper understanding of human teams is required. Distributed Cognition (DC) is a theoretical framework used to study how groups of humans cooperate. It views cognition as not confined solely to an individual's mind but distributed across individuals, artifacts, and tools in the environment. This short paper applies DC to the study of human-AI cooperation by analysing cooperative practices in two boardgames of particular interest to the AI community. Using DC we show how humans utilise the physical environment to communicate and perform cognitive work. We find that the presence of an AI assistant interferes with mental model formation and theory of mind reasoning, and we discuss potential causes for this interference through the lens of DC. Finally, we discuss how DC could be used in future work to understand how cooperation is achieved and what challenges arise in human-AI teams.

CCS Concepts: • **Human-centered computing → HCI theory, concepts and models**.

Additional Key Words and Phrases: Human - AI teaming, Theory of Mind, Distributed Cognition

ACM Reference Format:
Matthew Sidji, Wally Smith, and Melissa J. Rogerson. 2018. Adopting the Theory of Distributed Cognition for Human-AI Cooperation. In *Proceedings of Nov 30–Dec 4, 2024 (OzCHI'24).* ACM, New York, NY, USA, 11 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Currently there is significant interest in human - AI cooperation, where AI acts more as a teammate or peer to humans rather than being used merely as a tool. Yet, achieving successful human-AI cooperation and teaming has proven complex. Well-functioning teams carry with them unique social and cognitive challenges such as coordinating around common goals, forming shared mental models, theory of mind reasoning and higher needs for trust [15, 38]. Since AI agents in these settings will be affected by these cognitive aspects of teamwork, designers of these systems would benefit from a solid understanding of how cognitive work is being done by individuals in a team and by the team as a whole. In order to achieve this, we require a deep understanding of how human teams communicate and reason and how these cooperative process are supported or impeded by technology, AI or otherwise.

The complexity of successful human-AI teaming is exacerbated when teaming is required in scenarios with limited communication, and partial information. Despite their complexity, these scenarios are often compelling places to include human-AI cooperation as the potential benefit of including well functioning AI teammates is significant. Some real world examples include command and control scenarios, self driving vehicles on the road and co-creation tasks [4, 23, 37]. Studying and testing human-AI cooperation in these scenarios can be difficult as the cost of errors is often high (as in self-driving or command and control scenarios). Because of this, Partial Information, Restricted Communication, Cooperative (PIRCC) games have been identified as an ideal setting in which to study human-AI teaming [16, 34].

Recent approaches to study the cognitive practices of human-AI teams in PIRCC settings have often taken an approach to cognition where they model each agent's cognition internally [8]. In this model, cognitive events such as the manipulation of symbols or patterns of activation take place inside individual actors [21]. Human-AI cooperation research has been particularly interested in one important cognitive aspect in cooperation, theory of mind reasoning. Theory of Mind (ToM) reasoning is an agent's ability to reason about its own and others' mental states, including intentions, beliefs, and desires [13]. Imparting AI agents with ToM reasoning abilities often involves modelling nested internal mental models of teammates, and assuming others are doing the same [28] . Yet, what this approach fails to account for is that the cognition of a team often goes beyond just the collection of the internal minds of the individuals within the team [21]. This means that these internal cognitive theories often lack the tools to analyse how technology, embodiment and social structure contribute to groups solving problems, which we know are important factors in successful teams [20]. This is where the theory of Distributed Cognition can be helpful.

Distributed Cognition (DC) holds that the structures and processes of a cognitive system are distributed between internal and external representations, across groups of individuals, and across space and time [20]. The theory gained popularity in HCI and CSCW spaces as it emphasised the importance of the context and environment in which cognitive work is being done [36]. This approach better captures the reality of where and how technology is being used, and has been applied to understand how teams of humans cooperate [22, 27]. In this paper, we explore the potential of using DC as a theoretical basis for understanding ToM reasoning in human-AI cooperation. We do this by first exploring relevant literature then discussing two human-AI cooperation studies we conducted using the lens of DC. Following this, we expand on how these and further aspects of DC should be utilised in the study of human-AI cooperation and ToM reasoning.

2 RELATED WORK

In this section we will cover DC in more detail. We will then look into human-AI interaction research that has drawn upon aspects of DC. Finally we expand on the importance and use of theory of mind reasoning in human-AI cooperation research.

2.1 Distributed Cognition

Distributed Cognition, like other cognitive frameworks, sees cognitive processes as those involved with memory, decision making, inference, reasoning, and learning. Yet, unlike other cognitive theories it differs in where it recognises the boundaries of the unit of analysis and the range of mechanisms that participate in the cognitive process [21]. The theory focuses on three distributions of cognitive processes. Firstly, externalised cognition which involves the coordination between internal and external structures. For example how do physical objects in the environment support and enable cognition? Secondly, temporally distributed cognition, which are cognitive operations distributed through time. For example, how does cognitive work done in the past support current cognitive processes. Finally, socially distributed cognition, which is how cognition is supported by the social structures of a group. In his work, Hutchins advocates for "investigating the origins of symbols at the boundaries of our various units of analysis". This would involve how symbolic representations of cognition change when moving through a group, between individuals, the environment they are located, and through the technology they use to support cognition [21].

In the context of HCI's waves, Distributed Cognition was adopted in second wave HCI, where scope was broadened from the first wave's focus on usability and human factors to include cognitive theories, and a focus on group work in productivity settings [7]. DC was a suitable framework for investigating these issues and settings due to its focus on groups, and how artefacts can support cooperative work. It has been used to understand how pilots, naval officers, and software engineers cooperate and are supported by technology [11, 20, 22]. The theory has proven useful in understanding how complex team structures cooperate and how this is supported by technology. This therefore shows promise as a fruitful avenue for the study of human-AI cooperation.

2.2 Distributed Cognition in Human-AI interaction

Recent work has used DC to understand aspects of human-AI interaction more broadly. Aarset and Johannessen reflect on the process of developing a personalised AI textbook for a learner. Using DC they better understand the importance of the whole learning environment, beyond just considerations of the learner and teacher [1]. Grinschgl and Neubauer explore the potential for using AI to distribute cognition via cognitive offloading. They outline the benefits and harms of using AI to externalise cognition [14]. Naikar et al. apply the frame work of DC to inform the design of distributed systems involving human-autonomy teaming [31]. While these works utilise DC in understanding aspects and applications to human-AI interaction, there is little work utilising DC to understand human-AI interaction in actual cooperative scenarios.

2.3 Theory of Mind in Human - AI cooperation

An important aspect of human's ability to cooperate is our use of theory of mind (ToM) reasoning. ToM reasoning is an agents ability to reason about their own and others' mental states, which includes intentions, beliefs, and desires [29]. Achieving human levels of ToM reasoning in AI has become an important goal for researchers in human - AI cooperation [38]. Much of the work on ToM in human - AI cooperation has taken an internal or computational approach to modelling human cognition [9, 28]. While this approach more closely models the internal architecture of AI, we know from Hutchins (1995) that the cognitive practises within groups of humans cannot be solely explained by only considering individual agents [20]. The internal approach to cognition has also lead to difficulty in identifying the aspect of human cooperation that lead them to outperforming AI in similar tasks [35]. Given the importance of DC for understanding how teams of humans operate, and considering that cooperative settings frequently facilitate ToM reasoning, we see potential in using DC as a lens to understand human-AI cooperation and in particular ToM reasoning.

3 CRITICAL REFLECTIONS

In this section we critically reflect on two studies that investigated the cooperative practises of human and human - AI teams. Through using the theory and language of DC we reveal the important cognitive processes that occur which reveal how human teams operate, and the challenges faced when introducing AI into existing cooperative groups. Each section details the aim of the study, the context of the cooperative task, and reflects on the findings through the lens of DC. Specifically, these studies discuss two sub-topics within DC. Firstly, what happens to ToM reasoning when working with teammates that are distributing cognitive processes between human and AI agents? Secondly, how do humans utilise externalised cognition in order to perform theory of mind reasoning with strangers?

Fig. 1. Codenames word diagram. Red and blue teams words are highlighted. The card in the bottom left contains the secret key, indicating to the Cluegivers which words to clue for.

3.1 Codenames: Theory of Mind Reasoning of Hybrid Intelligence

In this study we were interested in how Large Language Model (LLM) assistants affect communication, reasoning, and social dynamics during the game *Codenames*.

3.1.1 Study Set Up. *Codenames* is a cooperative, restricted-communication, word association game in which players give clues to allow their teammates to guess secret words on the table. Players are split into two teams, with one Cluegiver and multiple Guessers on each. Each turn the Cluegiver gives a one word clue to their Guessers to try and lead them into selecting their teams secret words on the table. Each team takes turns giving clues and guessing until one team guesses all their words (see Figure 1). The game requires language-based reasoning, but also ToM as guessers often use what they know about Cluegivers to decipher the meaning of their clues. An important aspect to playing the game well is the ability for teams to form mental models. We conducted ten in-person *Codenames* sessions with a total of 54 participants. Half of the participants played in groups with their friends, while the other half played with strangers. Each session contained two teams of two to three players for a total of four to six participants each session. A session consisted of three games of *Codenames* followed by a 30 minute focus group. In each game, the Cluegiver had access to the AI clue-giving agent on a laptop in front of them. Cluegivers were asked to use the AI to generate clue suggestions for at least the first three turns in each game but were not required to give those suggestions as clues. After the third turn, Cluegivers were free to stop using the AI if they wished. It should be noted that all but one participant chose to continually engage with the AI after the initial 3 rounds. After each round, the Cluegiver was asked to swap with another member of their team so all participants had at least one game as Cluegiver with the AI agent. Focus group data was transcribed and analysed using Braun and Clark's reflexive thematic analysis with a focus on the cognition, social dynamics and player experience [5].

The decision to provide the Cluegiver with the LLM assistant was made as we were interested in the AI's effect on hidden information in PIRCC games. In Codenames there is no restriction on Guessers' communication, only the Cluegiver. We therefore chose to supply the Cluegiver with the AI assistant as this would be more applicable to understanding AI's effect in other PIRCC settings.

3.1.2 Reflection on Results. One interesting takeaway from this study was that performing ToM reasoning on Cluegivers when they were using an AI assistant was exceedingly difficult for Guessers as they had to now take into account that they were dealing with a hybrid intelligence. By hybrid intelligence we mean that the Cluegiver's cognition was now distributed between the human Cluegiver and their assistive AI. Guessers reported having to apply a mental filter on clues to assess how much was influenced by the AI and how much by the human. Participants reported this being a difficult and frustrating step, often leading to confusion and miscommunications. Although the LLM agent could only provide suggestion to the Cluegiver, its social presence as "another player" was felt by many participants throughout the games. We observed players referring to the AI using pet names, considering if clues "belonged" to the AI, and commenting on how they had to take its style into account as much as they would with any human Cluegiver. The LLM agent even interfered with ToM reasoning between players who had an existing relationship, had played together before and knew each other well. This suggests that performing ToM reasoning on a hybrid intelligence is increasingly difficult even when the humans already know each other and have worked together in the setting before. What made this reasoning especially difficult was the fact that Cluegivers could not reveal to Guessers if a clue was coming from the AI or themselves. Most of the time, the clue was generated by both the human and AI, as the human would alter or partially use an AI suggested clue. In this case the role of Cluegiver can be viewed as a form of distributed cognition, where the cognitive processes of the clue-giving was distributed between themselves and the AI. Guessers reported that when they were performing ToM reasoning on the Cluegivers, they attempted to form mental models of human Cluegiver and AI assistant to human Cluegiver. Evidence for this came from Guessers reporting applying an "AI filter" to given clues to check the clue against their understanding of how AI might give clues, as well as participants reporting that they would take into account how similar the human Cluegiver's style was to that of the AI. Further complexity arises through the imbalanced power dynamic between Cluegiver and AI, as ultimately the human has the final decision on what clue to give. For the Guesser to know how much of the cognitive process should be attributed to the AI, they must understand the human's beliefs about the AI. These beliefs could be viewed as a third mental model the Guessers must develop, the human's mental model of the LLM agent. As attempts to implement human - AI collaboration become increasingly popular, we see it likely that these hybrid team arrangements – where humans must perform ToM reasoning on their hybrid human-AI teammates – will appear frequently in the future.

3.2 Hanabi: Theory of Mind Reasoning with External Cognition

In another study, we were interested in the cognitive practises used by players of the hidden information game *Hanabi*, which helped explain humans superior performance to AI. In *Hanabi* players need to embed implicit information into clues they give to teammates, which requires the use of ToM reasoning to decipher them. Given that *Hanabi* requires ToM to be played well, AI researchers are using it as a test bed for developing agents with ToM reasoning [2]. At present, these agents perform far worse than human players [18]. Surprisingly, we found that there was little research into what humans were doing that allowed even first time players to play *Hanabi* so effectively. Our study aimed to identify the communication and reasoning techniques players were using that allowed them to play successfully [35].

Fig. 2. Set up of a Hanabi game. The player's cards face outwards so that the other players can see them.

3.2.1 Study Set Up. *Hanabi* is a game for 2 - 5 players in which they must share information in order to create stacks of ascending cards [3]. There are five different "suits" of cards, with cards numbered one to five in each. A player cannot see the cards in their own hand as they are held facing outwards (see Figure 2). On each player's turn, they must choose one of three actions: A player can play a card from their hand, discard a card from their hand, or give a clue. A clue is given to a single teammate, about the cards in their hand. The clue can only tell a teammate about the colour or number of specific cards, and must reveal all cards in that players hand with the clued colour or number. Outside of clues players are not allowed to communicate with one another. There are a limited number of clues that can be given each game, represented by the clue tokens. Once there are no clue tokens left players must discard cards from their hand to get clue tokens back. If a player plays the correct card it goes on top of or starts a new pile. If the card placed is incorrect, the players lose one of their three "lives". The game is over when all 25 cards have been placed, the deck runs out, or all lives are lost.

In this study, we observed *Hanabi* gameplay through a combination of overhead cameras and a first-person perspective view of participants using recordings of footage from eye-tracking glasses. We transcribed and the annotated sessions, and analysed our observations using thematic analysis. We found that the physical artefacts (pieces on the board) were essential to how players communicated aspects of their internal minds. Players externalised their intentions, beliefs and desires by rotating, reordering, and reconfiguring the cards in their hand. This created side channels of communication which other players used to perform ToM reasoning about their teammates. For example, players would rotate their cards in order to remind themselves, and communicate to others, which cards they had - or believed they had - information about (beliefs). Players would also use certain hand positions to indicate an intention to play or discard certain cards when it came back to their turn.

3.2.2 Reflection on Results. Here, the players externalised their cognition in order to reduce cognitive load. This process was also a means of making private information - their intentions, beliefs, and desires - publicly accessible. In this way, what was previously conceived as private information becomes publicly accessible because it is situated in the observable environment. We argue, that due to the importance of the physical artefacts for human ToM reasoning, AI agents should find ways to meaningfully interact with the externalised aspects of the game. This emphasises the importance of externalisation in the process of ToM reasoning and to human - AI cooperation more widely.

4 UTILISING DISTRIBUTED COGNITION FOR THEORY OF MIND REASONING IN HUMAN - AI COOPERATION

Stemming from findings in the studies presented we discuss applying DC to ToM reasoning in human - AI cooperation. Firstly, we discuss how teams externalise cognitive elements important to ToM reasoning, such as beliefs, intentions, and desires and how these can be further investigated under the lens of DC. Secondly, we explore how DC can help us better understand and design for teams made up of hybrid human-AI teammates. Finally we contextualise our approach within the various waves of HCI, showing how DC used in this way bridges the gap between 2nd, 3rd and 4th wave HCI.

4.1 Externalised Cognition

As detailed in 3.2 we see significant benefit in human - AI cooperation research focusing on the elements of a cooperative cognitive system. In cooperative scenarios with limited communication, humans will often externalise their cognition and utilise the physical environment as a cognitive aid, hence creating the observed side channels of communication. Through the act of externalising cognition the team members, whether intentionally or not, make aspects of their internal minds publicly observable by the wider team. If other team members are aware of what these externalised cognitive artefacts represent and how they are being used, teammates can use this to reason about each other. We therefore see great opportunity for cooperative AI to engage with this wealth of communicative information located in the physical environment.

Researchers might begin by observing human cooperative practices in context. What artefacts are utilised, what types of communication do they afford and how are these externalisations taught, interpreted and communicated? We also see potential in more design focused work in creating and testing these physical cognitive objects in environment which might benefit from them. This could allow the contents of the mind relevant to ToM reasoning - intentions, beliefs and desires – to be physically situated and hence observable by embodied cooperative AI. Due to the importance of the physical environment in this mode of DC we see this area of investigation being particularly relevant for human - robot interaction studies. Current work already focuses on how body language, gesture, and embodied communication affect human - robot interactions [24, 33]. It would seem to be a natural extension of this existing research direction to consider how physical elements of the environment can be utilised for robots to communicate and reason about human teammates, or vice-versa.

4.2 Hybrid Agent Cognition

The human agent using an assistive AI agent is an interesting and difficult cognitive element to conceptualise. As seen in our Codenames study, this team configuration can cause problems for team ToM reasoning as well as the formation of mental models. Consider the scenario where teammate A is using an AI agent and teammate B is attempting ToM reasoning on person A. Now let us explore two potential ways, within DC, that we could view this scenario. One way we could do this is to consider these as separate elements in the cognitive system. i.e. we should conceive of this as a human A using a cognitive tool, akin to writing on paper or using a calculator. Yet, this creates problems. Due to the non-deterministic nature of AI tools, especially LLMs, it is difficult for person B to form an accurate mental model of what possible outputs the AI can produce and therefore how this may affect person A's thinking. This view also requires a strong understanding from person B about what person A thinks about the AI, an example of second order theory of mind reasoning [30]. A second way we could envision this scenario is that the human - AI assistant dyad is a single hybrid mind. Instead of trying to separate the AI assistant from the human operator, and performing ToM reasoning on each, they could be viewed as a single cognitive unit within the team. In doing this a teammate might no longer attempt to reason over separate mind while performing ToM reasoning, but merely envision the dyad as a single system on which ToM reasoning can be performed.

From interviews with our Codenames participants, they seemed to take the former view. This was shown by the fact that many participants reported trying to conceive of what person A would think, what the AI would produce, and what person A thought about the AI, in separate steps. They were attempting to distinguish the two elements (person A and AI) and their relationship to one another. As we saw from our interview, this resulted in frustration, confusion and difficulty in interpreting clues. Most interestingly, there seemed to be no reported improvement in ToM reasoning when the teammates had played multiple times with one another. In this case we would expect person B to have an existing mental model of person A resulting in an easier time performing ToM reasoning. This expectation was supported by the reports from participants saying they typically understand their friends clues quite well. Yet, the existing familiarity with how their friend reasons in Codenames did not seem to improve ToM reasoning when the friend was using the assistive AI. This warrants further investigation as to whether existing mental models interfere or assist in ToM reasoning on hybrid teammates, as well as how organisational norms and structures could facilitate hybrid teammates. As illustrated in [10] this kind of hybrid team will be a common organisational structure going forward, and even now, as coworkers and teammates use cognitive AI tools such as ChatGPT [32]. This highlights the importance of thinking about how to best design AI assistants such that their output can aid, rather than interfere in humans' ToM reasoning.

4.3 Situating Distributed Cognition Across HCI's Waves

While Distributed Cognition (DC) can be considered a second wave HCI theory, it presents distinct features that make it compatible with the other HCI paradigms. Despite sharing a focus on cognition with other theories in first wave HCI, which are concerned with "the relationship between what is in the computer and the human mind" [17], DC does not strictly differentiate between the internal human mind and external cognitive tools. Instead, DC considers cognition as distributed across people, artifacts, and environmental elements within a system, dissolving the boundary between internal and external cognitive processes. This consideration of the external environment tied DC closely to second wave HCI's focus on workspace. Third wave HCI is defined by (among other things) its concern with user experience, non-work contexts, and centering the importance of embodiment and social context in which interaction occurs[6]. DC too shares a concern with how social and cultural factors shape and facilitate cognition [20]. This perspective allows DC to address both traditional productivity-focused interactions as well as socially and culturally embedded non-productive activities, which become the focus in 3rd wave paradigms. In particular, the settings of games and play become a suitable place to apply DC.

In this study, DC's flexibility facilitates its application within a 3rd wave framework, where the focus includes cultural, social, and non-work settings rather than solely on productive tasks. Through applying an observational method on naturalistic play behaviour, our research explores how cognitive processes are distributed among human players, AI agents, and game elements within the game setting. Although the studies are conducted in a lab, they retain a naturalistic character as players enter what Huizinga (1938) famously termed the magic circle—an invisible boundary marking the play space and distinguishing it from ordinary life [19]. Here players commit to the norms and rules of the game, creating a social environment around an authentic cultural activity rather than entering into an artificial test setting. Because of this, DC becomes a suitable framework to study human-AI cooperation in games due to the magic circle maintaining the 'in the wild' nature of the setting.

Moreover, this study's integration of DC aligns with emerging fourth wave HCI perspectives, particularly those centered on entanglement and post-humanist approaches, as outlined by Frauenberger in [12]. Like Actor-Network Theory (ANT), which broadens the notion of agency to include both human and nonhuman actors within an interaction network, DC inherently decentralizes human agency, recognizing nonhuman actors (in this case, AI agents and in-game artefacts) as integral contributors to the cognitive system [26]. The explicit connection to DC has also been noted by foundational scholars in ANT [25]. This post-humanist perspective is essential to entanglement HCI, where the focus shifts from human-centered design to relational ontologies, examining how various actors are interconnected within an ecosystem. By viewing artefacts, AI, and hybrid human-AI dyads as active agents within the gameplay system, this research underscores how DC can be used within 4th wave entanglement HCI to study interactions that are not exclusively human-centered, bridging human and AI roles in a network of distributed cognition.

5 CONCLUSION

Through these reflections on our two previous studies, we show the potential of human - AI cooperation research including a focus on Distributed Cognition. Adopting DC would align human - AI cooperation research more closely with the real-world functioning and cognitive processes of human teams. In particular, there is a need to address humans' utilisation of physical artefacts for externalising aspects of cognition, and the associated reasoning processes. We also contextualise our approach within the various paradigms of HCI, demonstrating how DC, a second wave theory, can be suitable for 3rd and 4th wave settings and problems. In addition, we recommend a focus on cooperative groups made up of hybrid human-AI teammates. By moving beyond the abstract, isolated computational models of the mind and highlighting the role of humans as embodied agents, we lay a crucial foundation for improving human - AI cooperation.

ACKNOWLEDGMENTS

Matthew Sidji's work is supported by the Australian Government Research Training Program Scholarship, by the University of Melbourne Faculty of Engineering and Information Technology Ingenium Scholarship Award, and by the LMA STELaRLab top up Scholarship.

REFERENCES

[1] Magne V. Aarset and Leiv Kåre Johannessen. 2022. On Distributed Cognition While Designing an AI System for Adapted Learning. *Frontiers in Artificial Intelligence* 5 (2022). https://doi.org/10.3389/frai.2022.910630

[2] Nolan Bard, Jakob N. Foerster, Sarath Chandar, Neil Burch, Marc Lanctot, H. Francis Song, Emilio Parisotto, Vincent Dumoulin, Subhodeep Moitra, Edward Hughes, Iain Dunning, Shibl Mourad, Hugo Larochelle, Marc G. Bellemare, and Michael Bowling. 2020. The Hanabi Challenge: A New Frontier for AI Research. *Artificial Intelligence* 280 (March 2020), 103216. https://doi.org/10.1016/j.artint.2019.103216

[3] Antoine Bauza. 2010. *Hanabi.* R&R Games.

[4] Francesco Biondi, Ignacio Alvarez, and Kyeong-Ah Jeong. 2019. Human–Vehicle Cooperation in Automated Driving: A Multidisciplinary Review and Appraisal. *International Journal of Human–Computer Interaction* 35, 11 (July 2019), 932–946. https://doi.org/10.1080/10447318.2018.1561792

[5] Virginia Braun and Victoria Clarke. 2012. Thematic Analysis. In *APA Handbook of Research Methods in Psychology, Vol 2: Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological.* American Psychological Association, Washington, DC, US, 57–71. https://doi.org/10.1037/13620-004

[6] Susanne Bødker. [n. d.]. Third-Wave HCI, 10 Years Later—Participation and Sharing. 22, 5 ([n. d.]), 24–31.

[7] Susanne Bødker. [n. d.]. When Second Wave HCI Meets Third Wave Challenges. In *Proceedings of the 4th Nordic Conference on Human-computer Interaction: Changing Roles* (New York, NY, USA, 2006-10-14) *(NordiCHI '06)*. Association for Computing Machinery, 1–8. https://doi.org/10.1145/1182475.1182476

[8] Tathagata Chakraborti, Subbarao Kambhampati, Matthias Scheutz, and Yu Zhang. 2017. AI Challenges in Human-Robot Cognitive Teaming. https://arxiv.org/abs/1707.04775v2.

[9] Tathagata Chakraborti, Subbarao Kambhampati, Matthias Scheutz, and Yu Zhang. 2017. AI Challenges in Human-Robot Cognitive Teaming. arXiv:1707.04775 [cs.AI]

[10] Allan Dafoe, Yoram Bachrach, Gillian Hadfield, Eric Horvitz, Kate Larson, and Thore Graepel. [n. d.]. Cooperative AI: Machines Must Learn to Find Common Ground. 593, 7857 ([n. d.]), 33–36.

[11] Nick V Flor and Edwin L Hutchins. 1991. Analyzing Distributed Cognition in Software Teams: A Case Study of Team Programming During. In *Empirical Studies of Programmers: Fourth Workshop*. Intellect Books, 36.

[12] Christopher Frauenberger. [n. d.]. Entanglement HCI The Next Wave? 27, 1 ([n. d.]), 2:1–2:27. https://doi.org/10.1145/3364998

[13] Douglas Frye, Philip David Zelazo, and Tibor Palfai. 1995. Theory of Mind and Rule-Based Reasoning. *Cognitive Development* 10, 4 (Oct. 1995), 483–527. https://doi.org/10.1016/0885-2014(95)90024-1

[14] Sandra Grinschgl and Aljoscha C. Neubauer. 2022. Supporting Cognition with Modern Technology: Distributed Cognition Today and in an AI-Enhanced Future. *Frontiers in Artificial Intelligence* 5 (2022). https://doi.org/10.3389/frai.2022.908261

[15] Victoria Groom and Clifford Nass. 2007. Can Robots Be Teammates?: Benchmarks in Human–Robot Teams. *Interaction Studies* 8, 3 (Dec. 2007), 483–500.

[16] Rotem D Guttman, Jessica Hammer, Erik Harpstead, and Carol J Smith. 2021. Play for Real (Ism)-Using Games to Predict Human-AI Interactions in the Real World. *Proceedings of the ACM on Human-Computer Interaction* 5, CHI PLAY (2021), 1–17.

[17] Steve Harrison, Deborah Tatar, and Phoebe Sengers. [n. d.]. The Three Paradigms of HCI. 1–18.

[18] Hengyuan Hu, Adam Lerer, Alex Peysakhovich, and Jakob Foerster. 2020. "Other-Play" for Zero-Shot Coordination. In *Proceedings of the 37th International Conference on Machine Learning*, Daumé Hal, III and Singh Aarti (Eds.), Vol. 119. PMLR, Proceedings of Machine Learning Research, 4399–4410.

[19] Johan Huizinga. [n. d.]. *Homo Ludens: A Study of the Play Element of Culture*. Beacon Press.

[20] Edwin Hutchins. 1995. *Cognition in the Wild*. The MIT Press, Cambridge, MA, US. xviii, 381 pages.

[21] Edwin Hutchins. 2000. Distributed Cognition. *International Encyclopedia of the Social and Behavioral Sciences. Elsevier Science* 138 (2000), 1–10.

[22] Edwin Hutchins and Tove Klausen. 1996. Distributed Cognition in an Airline Cockpit. *Cognition and communication at work* (1996), 15–34.

[23] Angel Hsing-Chi Hwang and Andrea Stevenson Won. 2021. IdeaBot: Investigating Social Facilitation in Human-Machine Team Creativity. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3411764.3445270

[24] Aleksandra Kalinowska, Patrick M. Pilarski, and Todd D. Murphey. 2023. Embodied Communication: How Robots and People Communicate Through Physical Interaction. *Annual Review of Control, Robotics, and Autonomous Systems* 6, Volume 6, 2023 (May 2023), 205–232. https://doi.org/10.1146/annurev-control-070122-102501

[25] Bruno Latour. [n. d.]. Cogito Ergo Sumus! Or Psychology Swept inside out by the Fresh Air of the Upper Deck [Review of Hutchins' Cognition in the Wild, MIT Press, 1995]. 3, 1 ([n. d.]), 54–63.

[26] Bruno Latour. [n. d.]. *Reassembling the Social: An Introduction to Actor-Network-Theory*. OUP Oxford. googlebooks:BgJREAAAQBAJ

[27] Reeva Lederman and Robert B Johnston. 2011. Decision Support or Support for Situated Choice: Lessons for System Design from Effective Manual Systems. *European Journal of Information Systems* 20, 5 (Sept. 2011), 510–528. https://doi.org/10.1057/ejis.2011.11

[28] M. Eger, C. Martens, P. S. Chacón, M. A. Córdoba, and J. Hidalgo-Cespedes. 2021. Operationalizing Intentionality to Play Hanabi With Human Players. *IEEE Transactions on Games* 13, 4 (Dec. 2021), 388–397. https://doi.org/10.1109/TG.2020.3009359

[29] Eric Margolis, Richard Samuels, and Stephen P Stich. 2012. *The Oxford Handbook of Philosophy of Cognitive Science*. Oxford University Press, USA.

[30] Ben Meijering, Hedderik Van Rijn, Niels Taatgen, and Rineke Verbrugge. 2011. I Do Know What You Think I Think: Second-order Theory of Mind in Strategic Games Is Not That Difficult. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 33.

[31] Neelam Naikar, Ashleigh Brady, Glenn Moy, and Hing-Wah Kwok. 2023. Designing Human-AI Systems for Complex Settings: Ideas from Distributed, Joint, and Self-Organising Perspectives of Sociotechnical Systems and Cognitive Work Analysis. *Ergonomics* 66, 11 (Nov. 2023), 1669–1694. https://doi.org/10.1080/00140139.2023.2281898

[32] OpenAI. [n. d.]. *ChatGPT: Optimizing Language Models for Dialogue*. https://openai.com/blog/chatgpt

[33] Sarah Schömbs, Saumya Pareek, Jorge Goncalves, and Wafa Johal. 2024. Robot-Assisted Decision-Making: Unveiling the Role of Uncertainty Visualisation and Embodiment. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–16. https://doi.org/10.1145/3613904.3642911

[34] Matthew Sidji, Wally Smith, and Melissa J. Rogerson. [n. d.]. Human-AI Collaboration in Cooperative Games: A Study of Playing Codenames with an LLM Assistant. 8 ([n. d.]), 1–25. Issue CHI PLAY. https://doi.org/10.1145/3677081

[35] Matthew Sidji, Wally Smith, and Melissa J. Rogerson. 2023. The Hidden Rules of Hanabi: How Humans Outperform AI Agents. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3544548.3581550

[36] Phil Turner. 2016. Post-Cognitive Interaction. In *HCI Redux: The Promise of Post-Cognitive Interaction*. Springer International Publishing, Cham, 75–76. https://doi.org/10.1007/978-3-319-42235-0_9

[37] Karel Van Den Bosch and Adelbert Bronkhorst. 2018. Human-AI Cooperation to Benefit Military Decision Making. NATO.

[38] Jessica Williams, Stephen M. Fiore, and Florian Jentsch. 2022. Supporting Artificial Social Intelligence With Theory of Mind. *Frontiers in Artificial Intelligence* 5 (Feb. 2022). https://doi.org/10.3389/frai.2022.750763