Towards Fluid Human-AI Collaboration

Fabian Heinrich, Florian Schröder, Stefan Kopp
Social Cognitive Systems Group,
CITEC, Bielefeld University
{fheinrich, fschroeder, skopp}@techfak.de

Abstract

In this presentation, we explore the concept of *fluid collaboration* (FC) which describes highly flexible, real-time teamwork that relies on dynamic task and role allocation, as commonly observed in everyday human collaboration. We argue that AI agents need to leverage both online Theory-of-Mind inference and natural language capabilities to support the required mix of implicit and explicit task coordination. We present the "Cooperative Cuisine" testbed–an extended Overcooked-like environment–allowing for exploring the design of cooperative agents and adaptive human-AI teams to achieve *FC*.

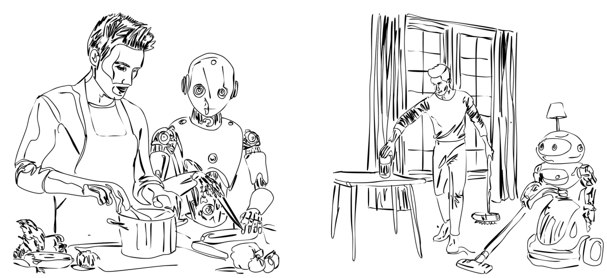

Figure 1: Examples of *fluid collaboration* in everyday tasks.

Introduction

Many tasks in everyday life require a myriad of abilities to be executed successfully. Surprisingly, collaborating with others often is easier and more efficient, even though interacting with them induces additional complexity and coordination load. For example, coordinating who grasps a cup to put it in the cupboard, or deciding who cuts the tomato and who cuts the cucumber require a complex reasoning chain that entails many details from individual abilities to experiences and preferences. What's more this collaboration is possible even in unstructured or dynamic environments. Our goal is to enable artificial agents to participate in these kinds of everyday collaborations with humans or other agents.

We adopt the term *fluid collaboration* (FC) to denote a distinct mode of collaboration that is flexible, adaptive, and does not follow strict protocols (cf. Ruvalcaba and Rogoff (2022)). This kind of collaboration can occur in almost all everyday life situations in which we interact with other people. Consider, for example, cleaning a room, preparing and serving a meal, or setting the table together. In all of these settings, there is no inherent fixed structure or objectively right way of doing things. Instead, who does what and in which order is determined dynamically and coordinated on-the-fly between the collaboration partners, either through proactive or reactive behaviors or through verbal communication and non-verbal cues.

We argue to study how *fluid collaboration* can be enabled between humans and AI-agents. This would significantly advance the AI and robot research community as it opens up new ways of cooperatively interacting with intelligent systems across various domains. Further, it implies tackling fundamentally hard research questions that need to be addressed in order to enhance the cooperative abilities of AI-agents and turn them into versatile, genuine collaborators.

Scenario and Approach

We operationalize FC using the *Cooperative Cuisine* environment (see Figure 2). It is built on the "Overcooked"-setting, a game in which players control chefs to prepare continuously ordered meals in a virtual kitchen environment. While already being frequently used in collaboration research (Bishop et al. 2020; Carroll et al. 2019; Wu et al. 2021), we extend the framework by enabling multiple humans and AI agents to collaborate in a highly situated manner on shared tasks and as potentially equal partners that have the same capabilities in the environment. To increase realism, we added continuous movement control and physical interaction (e.g. collisions or moving each other). Further, we extend the task space towards scalable real-life complexity while retaining experimental control.

We believe that *FC* can emerge naturally when humans and agents collaborate in this scenario. For one, the environment is highly dynamic and requires flexible coordination as orders come in in a fast-paced and unpredictable way. Thus, all collaboration partners need to be able to adapt quickly to the situation as well as to other agents. This flexibility allows the team to handle unexpected changes or newly emerging tasks. A defining characteristic of FC is hence that roles and responsibilities are not fixed in advance but evolve naturally during the interaction. We have seen this fluidity already in first empirical studies on human-human collab-

Figure 2: Screenshot of Cooperative Cuisine: A human actor and an AI agent prepare meals together.

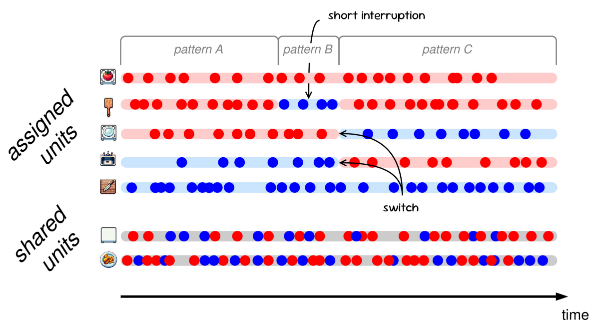

Figure 3: Illustration of fluid collaboration with dynamically adapted, shared and assigned tasks or resources (units).

oration in this environment, where the dynamic allocation of tasks and resources to be used exclusively or jointly can be observed (see Figure 3 for an illustration; see Schröder, Heinrich, and Kopp (2024) for further details). On a more general note, fluidity of human-AI collaboration in this environment can be assessed through multiple lenses. Objective performance metrics capture how effectively teams complete tasks, adapt to new conditions, and handle interruptions. Behavioral communication measures indicate explicit negotiation effort. Subjective feedback from human participants offered insights into the perceived team cohesion and usage of Theory of Mind reasoning, as well as trustworthiness, transparency, and overall enjoyment of the interaction. Combining these measures, we can comprehensively study FC and identify agent architectures that foster more fluid and effective collaboration.

Regarding advances in the development of AI-based agents, the Cooperative Cuisine environment poses important challenges. In fast and dynamic environments, not every aspect can be conveyed through verbal communication, as explaining every step in detail just may be too complex to convey fast enough. In such situations, humans are constantly observing what other persons are doing. From their actions, they can infer what the other is trying to accomplish in that moment, and more so, adjust one's own actions and plans accordingly. This so-called ability of Theory of Mind (ToM) or ToM-reasoning is crucial for *fluid collaboration* and the success of the team in Cooperative Cuisine.

Agents that employ ToM can proactively adapt their behavior by predicting what another actor is aiming to do, thereby reducing the need for explicit communication. Yet, misunderstandings or unexpected events still arise and the ability to switch to explicit, natural language dialogue becomes crucial for clarifying uncertainty or conflicts. By integrating both mechanisms for coordination–implicit inference and explicit explanation–agents can ensure smoother task progress with minimal disruptions. A key challenge here is to overcome the computational complexity of Theory of Mind (ToM), e.g., by adopting approaches of satisficing mentalizing (Pöppel and Kopp 2018) or action-driven mentalizing (Schröder and Kopp 2024).

Finally, Cooperative Cuisine creates a need for verbal communication to ensure mutual understanding so that humans can predict an agent's behavior and trust its decisions. We have started to imbue agents with (LLM-based) language-processing abilities, showing the range of important questions and benefits of linking perception, reasoning, acting and communication in FC. For example, well-timed explanations can increase dependability and facilitate smoother exchanges when challenging tasks arise.

References

Bishop, J.; Burgess, J.; Ramos, C.; Driggs, J. B.; Williams, T.; Tossell, C. C.; Phillips, E.; Shaw, T. H.; and Visser, E. J. d. 2020. CHAOPT: A Testbed for Evaluating Human-Autonomy Team Collaboration Using the Video Game Overcooked!2. In *2020 Systems and Information Engineering Design Symposium (SIEDS)*, 1–6.

Carroll, M.; Shah, R.; Ho, M. K.; Griffiths, T. L.; Seshia, S. A.; Abbeel, P.; and Dragan, A. 2019. On the Utility of Learning about Humans for Human-AI Coordination. *Advances in neural information processing systems*, 32.

Pöppel, J.; and Kopp, S. 2018. Satisficing models of bayesian theory of mind for explaining behavior of differently uncertain agents: Socially interactive agents track. In *Proceedings of the 17th international conference on autonomous agents and multiagent systems*, 470–478.

Ruvalcaba, O.; and Rogoff, B. 2022. Children's fluid collaboration versus managing individual agendas: Cultural differences in pair programming. *Journal of Applied Developmental Psychology*, 81: 101438.

Schröder, F.; Heinrich, F.; and Kopp, S. 2024. Dynamic Mentalizing in Fluid Human-Agent Collaboration. Submitted for publication to *Frontiers in Robotics and AI*.

Schröder, F.; and Kopp, S. 2024. Towards Action-Driven Mentalizing in Realtime Human-Agent Collaboration. In *Workshop on Theory of Mind in Human-AI Interaction*.

Wu, S. A.; Wang, R. E.; Evans, J. A.; Tenenbaum, J. B.; Parkes, D. C.; and Kleiman-Weiner, M. 2021. Too many cooks: Bayesian inference for coordinating multi-agent collaboration. *Topics in Cognitive Science*, 13(2): 414–432.