An Interaction Game Framework for the Investigation of Human–Agent Cooperation

Philipp Kulms, Nikita Mattar, Stefan Kopp

Social Cognitive Systems Group, Center of Excellence 'Cognitive Interaction Technology' (CITEC), Faculty of Technology, Bielefeld University, Germany
{pkulms, nmattar, skopp}@techfak.uni-bielefeld.de

Abstract. Success in human-agent interaction will to a large extent depend on the ability of the system to cooperate with humans over repeated tasks. It is not yet clear how cooperation between humans and virtual agents evolves and is interlinked with the attribution of qualities like trustworthiness or competence between the cooperation partners. To explore these questions, we present a new interaction game framework that is centered around a collaborative puzzle game and goes beyond commonly adopted scenarios like the Prisoner's dilemma. First results are presented at the conference.

1 Introduction

With advances in human–agent interaction and artificial intelligence, future interactions with technical systems are likely to be shaped like a *cooperation* between partners with complementary competencies [2]. In such teams, each agent has some degree of autonomy to handle dynamic situations and to make decisions within uncertain situations. One central question is: what does it require for a human to be willing and able to cooperate with a computer? We take further steps to investigate the potential of virtual agents to support cooperation. We focus in particular on the central factors *trustworthiness* and *competence* and how they develop in, and influence an ongoing human–agent cooperation. We consider trustworthiness as the agents ability to signal reliability in terms of benevolent intentions [3], and we seek to analyze how it is intertwined with the user's perception of the agent's competence.

2 Theoretical Background

Cooperative interaction moved into focus of HCI with a paradigm shift from using computers as mere tools, to interacting with them as intelligent and autonomous partners that resemble human-like counterparts [5, 2]. Hoc [4] argued that cooperative situations are shaped by a) both agents' ability to interfere with each other on goals, resources, etc. (thus requiring coordination) and by b) the management of interference, for example to support the other agents tasks. To further promote the development of cooperative agents, it is crucial to

understand when and how cooperation emerges in scenarios that are dynamic, extend over many interactions, and afford communication via various social signals. Thus far, cooperative behavior is primarily investigated using idealized cooperative games originating from behavioral game theory (BGT). However, cooperative games such as the Prisoner's dilemma, although well-established, are limited in scope and ecological validity when it comes to studying issues like competence and trustworthiness and how they evolve in a scenario with richer interaction. These game scenarios are on purpose kept very limited, with various variations and modifications applied, e.g. recasting the Prisoner's dilemma into an investment game, or allowing emotional feedback. Aside from making binary decisions (cooperate: yes/no) or allocating money between both players, the decision spaces and possibilities to interact in a meaningful way are confined. Although the BGT approach has yielded remarkable results, it is difficult to model and examine cooperative human–agent interactions with cooperative games alone in the long run because they cannot sufficiently model settings that comprise cooperative elements of HCI such as the formation of teams, solving a collaborative task, or rich communication. We argue that a framework that allows for modeling and studying cooperation in human–agent interaction in a more realistic and valid, yet controllable and manageable way is still lacking. From our point of view, such a framework has to meet at least the following requirements: (R1) goal interdependence, (R2) evocation and observation of various key factors (e.g. trust/trustworthiness, cooperativeness, competence, willingness to take risks, believability, emotions), (R3) ability to communicate and coordinate between the agent, (R4) identification of utilities assigned to decisions, (R5) possibility to introduce agent embodiment, (R6) algorithmically feasible automatization, (R7) possibility to assign different roles to players.

3 An Interactive Cooperative Puzzle Game

We have developed an interaction framework that allows us to operationalize, manipulate and analyze the relevant factors and key characteristics of cooperative behavior in a systematic fashion (see Fig. 1). The general setting includes two players solving a puzzle game interactively. Inspired by Tetris®, the interaction scenario consists of a board where two players work together to place blocks of various shapes, using horizontal movements and rotation. In our game, blocks do not move down gradually and filled lines are not cleared. The former gives us the ability to manipulate the available time per round (difficulty), the latter simplifies the implementation of an algorithm that enables a virtual agent to participate as autonomous player. There are further reasons why we chose this puzzle game: First, it is relatively easy to implement a heuristic to find optimal solutions for an agent (R6). By adapting the underlying heuristic, we are thereby able to manipulate the agents behavior and thus, presumably, its perceived competence (R2). Second, we can induce both individual goals (e.g. as an individual score for a block a player has placed) as well as cooperative goals (as a team score for lines that get completely filled). This enables us to analyze if a player prioritizes
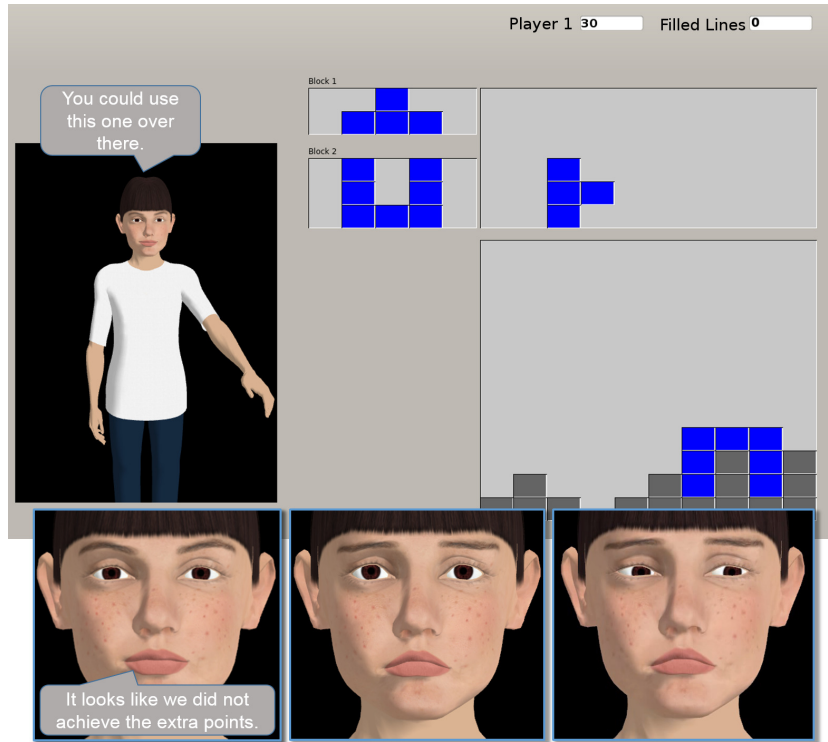
Fig. 1. Concept of the puzzle game interface with a virtual agent as cooperation partner. The agent can display multimodal behavior or other contextual social cues. This allows for analyzing which type of feedback cue (verbal, sad or regretful facial expression) has what differential effect on the ongoing cooperation.

one over the other as a measure of cooperativeness (R1). Third, by changing factors such as field size, time constraints, block shapes, or block scores we can manipulate the difficulty of achieving the cooperative goal. Finally, we can at all times fully observe the game state.

Two different player roles are introduced in an asymmetric setup of the game (R7): a *recommender* and a *decider*. Each round consists of the following steps: 1) The *recommender* picks one out of two blocks he wants to recommend to the *decider*. 2) The *decider* picks a block by either accepting or rejecting the recommendation and places it on the board. 3) The *recommender* places the remaining block. Both players receive points individually for every block they place. The complex block gives more points than the simple one, leading to an individual benefit in placing more complex blocks (R4). In addition, both players obtain points when a certain amount of lines are completely filled horizontally. We refer to this as cooperative or joint goal since it can only be achieved if the players cooperate with each other by recommending and choosing blocks that

minimize the occupied space, and by placing them in a reasonable way. The three actions occurring in this scenario – recommendations, (not) complying with recommendations, and placing blocks – relate to different behavioral factors (R2): recommendations and compliance are indicative of competence, trust, and pursued goals. For example, if the *recommender* constantly recommends the complex block this could be interpreted as a high level of trust in the partner, pursuit of an altruistic goal, but also as low level of competence of the *recommender* himself. In contrast, if the *recommender* always recommends the simple block this could be interpreted as low level of trust or pursuit of a selfish goal. Interpretations of the actions of the *decider* are quite similar. Still, one cannot always tell if the *decider* follows a recommendation or would have decided for the block anyway. The block placing may serve as indicator for level of competence, except for cases where players deliberately try to minimize their partners score by misplacing blocks. Note that in this framework, (non-)cooperation is strategic and requires more than a binary decision. Although the decision-making process in the often used social dilemmas is also complex, the final action translates to "I (do not) want to cooperate". In contrast, we expect the degree of cooperativeness the *decider* signals to the *recommender* to depend on a) the *decider's* puzzle solving skills, and b) the intention the *decider* signals by (not) complying with the *recommender's* advice. As a result, how is a *decider* perceived who performs well but only rarely accepts recommendations, or who accepts the recommendations but performs poorly? The perceived cooperativeness of the *decider*, then, should relate to the *recommender's* willingness to cooperate with him [1]. An experimental study where participants played as *recommender* and a non-embodied agent acted as *decider* showed that perceived competence, trustworthiness, and cooperativeness of the *decider* were differentially affected by its puzzle solving skills and the degree to which it accepted recommendations.

References

1. Balliet, D., Van Lange, P.A.: Trust, conflict, and cooperation: A meta-analysis. Psychological Bulletin 139(5), 1090 (2013)
2. Bradshaw, J.M., Dignum, V., Jonker, C., Sierhuis, M.: Human–agent–robot teamwork. IEEE Intelligent Systems 27(2), 2–7 (2012)
3. Deutsch, M.: Cooperation and trust: Some theoretical notes. In: Jones, M.R. (ed.) Nebraska Symposium on Motivation, pp. 275–320. University of Nebraska Press, Oxford, England (1962)
4. Hoc, J.M.: From human–machine interaction to human–machine cooperation. Ergonomics 43(7), 833–843 (2000)
5. Nass, C., Fogg, B., Moon, Y.: Can computers be teammates? International Journal of Human-Computer Studies 45(6), 669–678 (1996)