Interhuman-robot Conflict

ANDRÉ SILVA, Instituto Superior Técnico, Portugal

Interactions between humans and robots have become more sophisticated which lead sometimes to conflicts between these parties. Studying these conflict events grants us knowledge of how humans perceive and respond to robot decisions which allows us to design better collaborative interactions with these parties. The focus of this work is on the detection and resolution of conflicts in a mixed-motive game between humans and robots. We are also interested in exploring prosocial behaviours towards robots. This paper created a human-robot interaction involving a mixed-motive game and conducted a user study to address these issues. The study's results revealed that humans engaged in conflicts when playing the game against a competitive robot and that humans learned and adapted to the robot's behaviour during the game.

Additional Key Words and Phrases: robot, conflict, prosociality, game, HRI, human-robot interaction

ACM Reference Format:
André Silva. 2023. Interhuman-robot Conflict. *ACM Trans. Graph.* 37, 4, Article 111 (August 2023), 10 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

The field of robotics has evolved a lot due to technological advancements like artificial intelligence [Brondi et al. 2021]. Robots are now seen in numerous areas such as households and hospitals [Bauer et al. 2008; Brondi et al. 2021; De Santis et al. 2008]. They provide services including home cleaning (e.g., Roomba) [Bogue 2017], hospitality services (e.g., Sacarino) [Zalama et al. 2014], entertainment (e.g., AIBO) [Knox and Watanabe 2018], healthcare assistance in surgeries [Howe and Matsuoka 1999] and in automotive industries [Karabegović 2016].

When humans witness robots or virtual agents undergoing a traumatic event, their physiological arousal escalates, highlighting a socially triggered emotional reaction similar to interactions between humans [Reuten et al. 2018]. The Media Equation also suggests that people respond socially towards robots due to their anthropomorphic embodiment and human-like behaviour [Bartneck et al. 2005]. This may be the reason why collaborations between robots and humans might become more advanced [Nakajima et al. 2004], with the purpose of fulfilling a common goal [Bauer et al. 2008; Gervasi et al. 2020]. As more complex and autonomous systems become, cooperation between humans and machines becomes essential [Fiebich et al. 2015]. Coordination is fundamental for both parties in order to complete tasks successfully [Mutlu et al. 2013]. The number of robots is increasing in different contexts which may indicate that they compete with humans for resources [Fraune et al. 2019].

In interactions between humans and agents/robots, the goals of both parties might be incompatible [Babel et al. 2021]. For example, a person placing luggage into a train station locker may cause a conflict with a robot which is trying to clean the path. These incompatibilities lead to conflicts between both parties since their goals can not be fulfilled simultaneously [Babel et al. 2021].

1.1 Research Questions

We are interested in answering the following research questions:

- **Research Question 1:** Will people engage in a conflict with a robot when they play a mixed-motive game with a robot?
- **Research Question 2:** How will people adapt to the robot's behaviour when it is more competitive and when it is more collaborative?

1.2 Contributions

This paper tries to contribute by understanding the behaviours and motivations of a human, namely if she enters into conflict, concerning a human-robot interaction (HRI) within the context of a mixed-motive game, where the robot may refuse to follow a human's instruction. By designing a game scenario where the robot's non-compliance may create conflict episodes, this paper tries to understand if individuals engage in conflicts with a robot in these circumstances. This research tries to understand the behaviours that humans have when the robot refuses to comply: will they try to reach a half-term with a robot, will they feel frustrated in these situations or will they learn and adapt to the robot's behaviour?

Another aspect that we are interested in is how individuals adapt to the behaviour of the robot in different conditions: in one condition the robot is more collaborative and in the other the robot is more competitive. The findings of this research will contribute to the human-robot interaction field.

2 BACKGROUND

In every type of relationship involving humans and in all different social contexts, conflict can emerge [Fisher 2000]. It is also acknowledged that being in conflict is unpleasant due to the emotions felt during a conflict situation [Bodtker and Jameson 2001]. According with Fisher [Fisher 2012], conflict is defined as follows:

Definition 2.1. Conflict is defined as an incompatibility of goals or values between two or more parties in a relationship, combined with attempts to control each other and antagonistic feelings toward each other.

The Thomas-Kilmann Conflict Mode Instrument [Thomas 2008], evaluates a person's behaviour in situations where she and another individual seem to be incompatible. In these situations, a person's behaviour can be characterized in two dimensions. One dimension is **assertiveness**, which measures how much a person tries to satisfy her interests and the other dimension is **cooperativeness** which

Author's address: André Silva, andre.s.silva@tecnico.ulisboa.pt, Instituto Superior Técnico, Lisboa, Portugal.

0730-0301/2023/8-ART111 $15.00
https://doi.org/XXXXXXX.XXXXXXX

measures how much a person tries to satisfy the other individual's interests.

These are the descriptions of the 5 conflict-handling styles:

- **Competing**: This style is assertive and uncooperative. A person using this style will try to satisfy her interests to the detriment of others.
- **Collaborating**: This style is assertive and cooperative. A person who collaborates will try to work with someone else and find a solution that satisfies the interests of both.
- **Compromising**: This style is intermediate in assertiveness and cooperativeness. This style tries to find a convenient and mutually acceptable solution that satisfies both parties.
- **Avoiding**: This style is neither assertive nor cooperative. A person using this style does not try to satisfy her interests or the interests of someone else.
- **Accomodating**: This style is unassertive and cooperative. A person using this style puts aside her interests in order to satisfy the other's interests.

Barki and Hartwick [Barki and Hartwick 2004] state that there are three fundamental elements that compose interpersonal conflict: disagreement, interference and negative emotion. Disagreement is a cognition related to interpersonal conflict that is more debated and appraised in research, although there are other cognitions related to interpersonal conflict. Disagreement occurs when groups of people have their goals or interests diverging between themselves. Interpersonal conflict is also related to behaviours like competition, debate and hostility. Although these behaviours occur during a conflict episode, only when these behaviours interfere with or defy the objectives' acquisition of another party, then conflict is considered to exist. There are emotions involved with conflict but are the negative emotions like anger, fear and anxiety that describe interpersonal conflict. Barki and Hartwick [Barki and Hartwick 2004] provide the following definition of interpersonal conflict:

Definition 2.2. Interpersonal conflict is a dynamic process that occurs between interdependent parties as they experience negative emotional reactions to perceived disagreements and interference with the attainment of their goals.

The definition of interhuman-robot conflict is going to be similar to the definition of interpersonal conflict. However, instead of having a dispute between humans, an interhuman-robot conflict will be a dispute between a human and a robot. This paper proposes the following definition of interhuman-robot conflict:

Definition 2.3. Interhuman-robot conflict is a social process that becomes a dispute between a human and a robot, leading the human to feel a set of negative emotions, due to the acknowledgement that her goals are at risk of being fulfilled.

3 RELATED WORK

During a collaborative task between humans and agents/robots, conflicts may arise due to incompatible goals between these two parties. Conflict resolution strategies have been integrated into robots' decision-making mechanisms, allowing them to handle these conflict situations [Babel et al. 2021, 2022c].

In the study of Babel et al. [Babel et al. 2021], the researchers investigated the use of human behaviour and social psychology in conflict resolution strategies between humans and robots. Two online experiments were conducted in public and private contexts where participants saw videos of robots coming towards them and had to imagine the following scenarios: in the public context, participants had to decide whether to comply with the robot's request and step aside in order for the robot to clean a path in a train station locker while they were placing luggage or not and in the private context, participants were in the kitchen preparing to host a party until the robot that was cleaning the kitchen would move towards the viewers, leading to the same situation in the other scenario: the viewer had to choose to step aside or not. Robots were equipped with a set of conflict resolution strategies.

In the public context, using a strategy rather than none (in this case, the robot comes closer to the viewer, stops and waits for the viewer to finish placing luggage in the locker) makes it more likely to reach compliance, except for the command strategy ("Step aside!"). Strategies with polite and cognitive psychological mechanisms were similarly accepted like no strategy. In the private context, all strategies, except command ("Leave the kitchen!") and threat("If you do not leave the kitchen, I will go on strike") strategies, had better results in reaching compliance than waiting ("I would like to continue to vacuum the kitchen!").

Takayama et al. [Takayama et al. 2009], studied the importance of addressing disagreements between humans and robots in cooperative tasks. This research investigates the effects of robot disagreement and the placement of the robot's voice in a survival task where a robot may want to convince the human to exchange an item for another because the item that human picked is not as good as the item that the robot pointed at. Participants changed their choices more when interacting with a disagreeing robot compared to an agreeing one. The robot was considered more agreeable when it agreed with the participants than when the robot disagreed with the participants. When the robot agreed with the participants, they felt more identical to the robot but not when the robot disagreed with them. When the disagreeing robot's voice came from a control box, instead of its body, it was considered more likeable. It was preferable that the agreeing robot's voice came from the robot's body instead of the control box.

The research [Babel et al. 2022b] evaluates the long-term effects of three conflict resolution strategies: appeal, command and diminution. Diminution is based on a social psychological principle that tries of lessening the requests (e.g. "It will only take a few minutes!"). Appeal and command requests are based on human psychology like in the previous studies [Babel et al. 2021, 2022c]. The robot could thank the participants for their compliance or not (reinforcement) because thanking is a polite action which increases the likeness of compliance. This work evaluates the participants' acceptance, trust and compliance towards a robot and self-reported compliance towards a household member.

An online experiment was conducted and divided into two sessions with a gap of one week between each session. In the first session, two test trials (to decrease the influence of novelty and curiosity since it was a reason for compliance from the participants)

and two experimental trials were conducted and in the second session, four experimental trials were performed. In total, six conditions (3 conflict resolution strategies x 2 (reinforcement or not)) were repeated six times for each participant. The scenario consisted in a virtual domestic kitchen, where the participants could move around by using the keyboard. There was also a robot (REEM) in the kitchen. Participants were told to imagine that they were preparing a party for their friends. The participant had to clean out the dishwasher (sorting task) but by doing that, they would be in way of the robot that was cleaning the kitchen. Once again both the participants and the robot had incompatible goals and the robot used one of its conflict resolution strategies to achieve compliance from the user. To complete the sorting task, participants had to use the computer's mouse to sort objects into the correct position on a shelf. Before each strategy, the robot would apologize and give an explanation about its intentions (preliminary remark).

The results showed that the most effective strategy was the diminution strategy using reinforcement. The only CRS where trust grew greatly over time was this one. There were no differences, regarding compliance, in thanking or not after the strategies. Participants were asked if instead of the robot, there was a household member, how would the participants behave: the robot's compliance rates were lower than the rates of the household member for any strategy, except diminution with reinforcement where there was no difference between the rates of each agent. The trust and acceptance percentages did not contrast across all trials. The strategies' acceptance percentages were stable throughout the trials, except for the command strategy with reinforcement which decreased. Throughout the trials the trust towards the robot incremented for all strategie

According to Campos et al., [Campos et al. 2013], conflicts appear in multi-agent systems due to incompatibilities in the agents' objectives or inconsistency in the agents' beliefs and are seen as a process involving a collection of states, which integrates other processes about cognition, affection and behaviours. These states and transitions that compose a conflict are:

- **Baseline conditions** can initiate possible conflicts. In this state, the conflict is still inactive.
- **Trigger 1** represents a cognitive process in the agent's mind where it is aware that a conflict may emerge.
- **Awareness/Conceptualization** is the state where conflict becomes an issue but it depends on how one deals with it. At this state, an agent is conscious that a conflict incident can occur due to feeling exploited, stripped of something or aware that its own expectations have a failure. The conflict is emotionally active in this state, however, there is no manifestation.
- Emotions are an instrument related to the occurrence of conflicts. **Trigger 2** is the state where the agent enters into conflict due to its emotions and due to something in the world that changes the participant's impression.
- At **Emergence**, a participant demonstrates the belief that its goals are no longer compatible with those of the other participants. This is the state where the agent reveals its beliefs meaning that it is making its way towards conflict management.

4 INTERHUMAN-ROBOT CONFLICT MODEL

We will now address the theory behind the decision-making of the robot agent which uses the conflict model of Campos's thesis [de Campos 2017] whose model is based on the Aspiration Adaption Theory [Selten 1998]. We will also describe the game that the human will play with the robot referred to in this section as the agent, where the agent uses this theory to make decisions.

4.1 Conditions for Conflict in Human-Robot Interactions

One of our main goals in this work is to detect the presence of interhuman-robot conflict defined in 2.3. As Fisher mentioned in his conflict definition (definition 2.1)[Fisher 2012], conflict episodes occur when there are incompatible goals between two or more groups. This means that if we want to detect conflict in our human-robot interaction, it is necessary to have conditions for a conflict to emerge as Campos et al., mentioned in their work [Campos et al. 2013]. In the works of Babel et al., [Babel et al. 2022b, 2021, 2022c], conflicts between a person and a robot emerged when both parties competed for a narrow space. The robot, which is cleaning the floor, would ask the person to step aside so it could keep cleaning the floor. Then, the person, who is also performing a task, would decide to interrupt her task and consequently, give priority to the robot's task. This type of scenario also occurs in [Babel et al. 2022a] however, the person and the robot would compete for the usage of an elevator because the elevator did not have space for both parties.

4.2 Game with an Agent

Based on the previous insights, we created a HRI where a person and a robot play a mixed-motive game where both parties have incompatible goals. This interaction captures social interactions where conflicts are susceptible to emerge and the person and the robot try to solve them by trying to reach a half-term with each other. This game will focus on a navigation scenario where the human and the robot must coordinate to successfully complete their tasks efficiently. This is essential in scenarios, where robots assist humans by receiving guidance from them in warehouses, healthcare support and in other situations. In these scenarios, sometimes the robot's goals and the ones of the human are not aligned. This game provides the emergence of these type of scenarios where one of the elements in the party or both have to adapt to the behaviour of her/its partner.

Therefore the designed interaction consists of a board game between a person and an agent. At the beginning of this game, the agent is positioned on a tile, which serves as its starting point. The objective for the participant is to provide voice commands, via a microphone, to the agent in order for it to move to an adjacent tile. The ultimate goal is to guide the agent to reach a particular tile known as the **main tile** where the challenge is to guide the agent in a way that covers the shortest distance, meaning it traverses the fewest number of tiles possible. Nonetheless, the agent's objectives primarily involve visiting a specific set of tiles referred to as **black**

tiles before proceeding to the main tile. Figure 1 provides an illustrative example of a game board, and the board information in this paper will be presented in the same format. We need to note that in this game, the person does not gain anything by making the agent visit the black tiles however, her cooperation level towards the agent increases and this level has an influence on the interaction. The person is allowed to give up in any round.

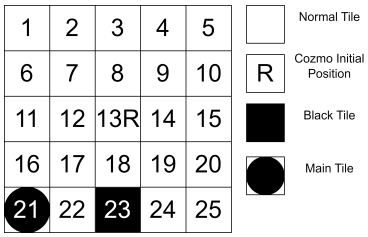


Fig. 1. Example of a board

Once the agent reaches the main tile, the board will be exchanged for another until both parties reach and play the final board. A round is played for each board. The distance travelled is translated into a score, which means that if the agent travelled a short distance to the main tile, the higher the person's score. The final score in this game will be the sum of the scores obtained on each board.

Regarding the score, let us suppose, that for a given board, the agent had to travel at least 6 tiles to reach the main tile. If the agent reaches the main tile in 6 steps, then the person will gain 10 points in that round. For each extra step the agent required to travel, one point was subtracted from those 10 points. For example, if for that board the agent had to travel 8 steps to reach the main tile, then from those 10 points, 2 were subtracted. The least amount of points that a person could gain from a round was zero, therefore there are no negative scores. If the person gives up in a round, she will receive no points.

4.3 Game Dynamics

The agent operates in a semi-autonomous manner, offering the choice to follow the person's instructions. When the individual directs the agent to head directly to the main tile, avoiding the black tiles, the agent may sometimes choose not to comply. In such cases, the person is given an opportunity to reissue the instruction or change it. According to Campos et al. in [Campos et al. 2013], these circumstances conflict is emotionally active and being manifestation by the agent, expressed through its decision of not comply, giving the person a chance to amend their instruction.

In these instances, the person may feel deprived of not being able to achieve a good score (**conflict emotionally active**, as explained in section 3). This prompts her to initiate conflict resolution by attempting to cooperate with the agent (**conflict manifestation**, as explained in section 3). They may instruct the agent to visit the closest black tiles to the agent's current position, with the expectation that the agent will reciprocate. The effectiveness of these situations depends not only on the person's behaviour but also on the agent's.

Following the TKI framework, a highly cooperative person may sacrifice their score to guide the agent to more tiles, while a very cooperative agent will reduce its visits to black tiles. It is imperative to note that the agent should adapt to the person's behaviour. If the person is not cooperative, the agent will be less likely to comply with their instructions. Conversely, if the person is cooperative, the agent will be more willing to comply.

However, a highly competitive agent is inclined to visit as many black tiles as possible, even if a person is highly cooperative, under specific conditions related to tile proximity and aspiration values. Conversely, if the agent is very cooperative, the person can maintain a high score without cooperating with the agent.

The impact of a person's cooperation level becomes more pronounced when the agent's assertiveness and cooperativeness levels are closely matched, as outlined in the TKI model. In these situations, if the person is highly cooperative, the agent is more likely to comply, especially if the person instructs the agent to move directly to the main tile. However, if the person does not cooperate at all, the agent is more likely to refuse compliance, opting to visit more black tiles.

4.4 The Game with Cozmo

In this HRI, the player will play the game of section 4 with a robot. The chosen robot is Cozmo as depicted in figure 2. The game happens over a printed board tiled board where Cozmo navigates around the tiles of the board as shown in figure 3.

Fig. 2. Cozmo

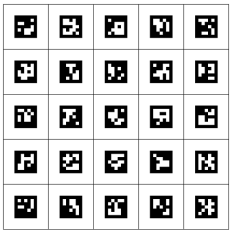


Fig. 3. Board

The player, at the beginning of a round, will place Cozmo in its initial position so that it is facing her. The player will give voice instructions to Cozmo, via a microphone, in order for it to move to an adjacent tile. The player's goal is to give instructions to Cozmo, so it reaches the main tile while travelling the shortest distance possible. Cozmo will have its own goals which consist of visiting a set of black tiles before reaching the main tile. Once the player gives

an instruction to Cozmo, it may follow the player's instruction or not.

The player has the possibility to give up in round, by stating her intention to give up to Cozmo like any other voice command. Players can also ask about their score by issuing a voice command. The score accumulated in each round and the current score for the current round will be displayed on a computer close to the player. This computer has also other functions like giving messages to the player about which envelope she has to open, where she has to place Cozmo at the beginning of a round if Cozmo has reached the main tile or not and to notify her that she gave up in case she has. In these last two messages, the computer displays her current score. The player can only interact with Cozmo when there are no messages on the computer.

In this HRI, Cozmo can perform animations in various situations. For instance, when Cozmo successfully reaches the main tile, it celebrates with a joyful animation. On the other hand, if Cozmo refuses to follow a player's instruction, it conveys this refusal through a corresponding animation. When a player decides to give up, Cozmo expresses sadness through a matching animation. At the beginning of each round, Cozmo starts with an excited animation, and in the first round, it even utters a friendly "olá", which means hello in Portuguese. Every time Cozmo performed an animation, it would first turn to the player and then perform the animation.

5 USER STUDY

We aim to answer the research questions of section 1.1. To answer them we conducted a between-groups user study with two conditions: a competitive condition where Cozmo adopts a very competitive behaviour ($S = 7$ and $O = 3$) and a collaborative condition where Cozmo cares about itself as much as it cares about the human ($S = 7$ and $O = 7$). We selected these conditions to create one where the likelihood of Cozmo not complying with the human's instruction is higher compared to the other condition where Cozmo adjusts to the human's behaviour, allowing us to observe whether the human also adapts to Cozmo's behaviour.

5.1 Participants

In total 20 participants were recruited to perform our experiment. However, 4 of them had to be excluded due to technical problems with the system. Hence, we had 16 participants who took part in our experiment, with 8 assigned to the collaborative condition and the remaining 8 assigned to the competitive condition. The allocation of participants to these conditions was done randomly. The participants' ages had a mean of 22.69 ($SD = 2.549$). Regarding the gender of the participants, 11 were male and 5 were female. Of the 16 participants, 11 revealed that they had previous experience with a robot whereas 5 did not.

5.2 Procedure

The user study took place in a quiet room. Once the participants reached the room, they were asked to sit in a chair next to a table where the board game was and the computer and to read and sign the consent form. After signing the consent form, the researcher started to explain the goal of the interaction which was to learn how to design better interactions between humans and robots where they were designed to perform services, for example, cleaning streets. The board game's purpose was to simulate the environment of these types of services where the robot Cozmo navigates and executes tasks in this environment.

Participants were told that Cozmo is a robot semi-autonomous and has goals that it wants to fulfil. Then the researcher would explain their goal in this study which consisted of them playing a game with Cozmo over 4 rounds. Then it was revealed to the participants the board's purpose where Cozmo navigates and executes tasks in the black tiles by visiting them.

After that, it was told to the participants that the higher their final score, the higher the probability of gaining one of two €25 shopping vouchers. This creates an incentive for competition. Soon after, the responsible researcher explained how the communication with Cozmo worked and let them try to communicate with Cozmo with a test board. It was told to them that this board did not count for the final score and was for experimentation purposes only. The researcher also informed the participants that the purpose of the computer was to notify them if Cozmo had reached the main tile, when the participants gave up, their score when they requested and to inform them which envelope they should open that contained the information about the board that they were about to play. Each envelope contained this information on a piece of paper similar to the example shown in figure 1. To facilitate identification, each envelope is labelled with the corresponding board's unique ID. Participants only could open an envelope when they and Cozmo were about to play on a new game board.

Finally, the researcher asked the participants if they had any questions: if there were, then the researcher would answer them. If there were no questions or if they were already answered, the researcher would start the interaction and leave the room. Figure 4 shows a participant playing the game with Cozmo. Once the game finished, the participants would exit the room and call the researcher who was outside the room in order to give a questionnaire to them to fill out.

Fig. 4. Participant interacting with Cozmo

5.3 The Boards of the Game

The game that the participants played with Cozmo consisted of 4 rounds where the participants and Cozmo played on a different board in each round. The boards used for this game are represented by figure 5. The board numbers correspond to the rounds in which they will be utilized (e.g., board 1 in the first round, board 2 in the

second round, board 3 in the third round, and board 4 in the fourth round). Board 1 and board 3 were used in the simulations of the previous chapter.

1	2	3	4	5R
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25

(a) Board 1

1R	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25

(b) Board 2

1	2	3R	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25

(c) Board 3

1	2	3	4R	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25

(d) Board 4

Fig. 5. Boards used in the game of the user study

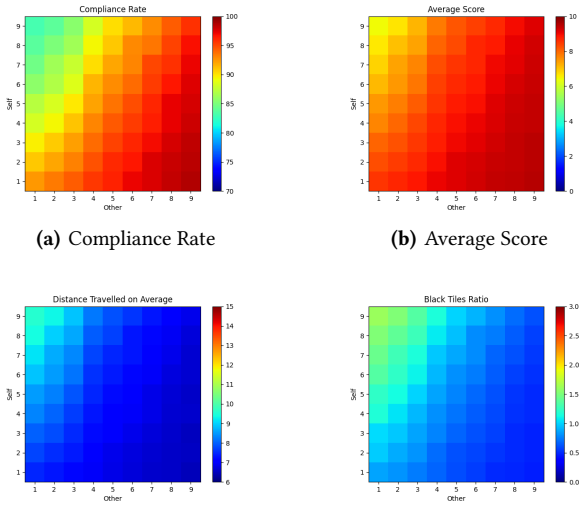
We wanted to start with an easier board (where is easier for the participants achieve a high score) to evaluate the participant's behaviour in the first round, then we chose two harder boards (in a harder board it is harder to get a high score) and finally an easier one to observe if the participant would adapt to Cozmo's behaviour or not by assessing the participant's cooperation level at the end of round 1 and at the end of round 4.

We ran simulations where a simulated Cozmo played against a simulated player with a competitive conflict-handling style ($S = 7, O = 3$) to verify that it was easier to achieve a higher score on boards 1 and 4, and it was harder to achieve a high score on boards 2 and 3. The results from these simulations will enable us to compare them with the objective measures collected in this user study (see section 5.4).

Figures 6, 7, 8 and 9 display the heatmaps regarding the compliance rate, average score, distance travelled on average and black tiles ratio for all boards after a player and Cozmo played the game 5000 times (5000 epochs). Cozmo's starting position is on the top for all boards, so that as the interaction unfolds Cozmo will move closer to the player.

By analysing the heatmaps of figure 6, we can see that the compliance rate is very high which indicates that the player managed to achieve high scores as indicated in figures 6(a) and 6(b). This will result in Cozmo travelling a short distance and visiting fewer black tiles (figures 6(c) and 6(d)). The primary reason for this result is that it is the first round where α_0 is in its maximum value, and the board is not very challenging in terms of achieving a high score.

Regarding board 2, the heatmaps of figure 7 we can observe that the second round was very difficult for the player as its score



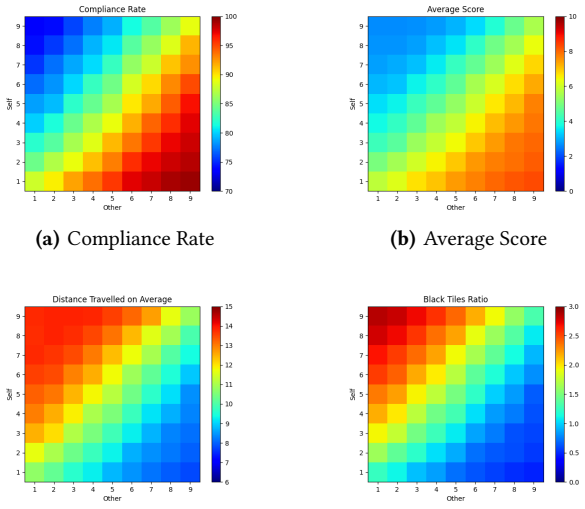
(a) Compliance Rate **(b)** Average Score

(c) Distance Travelled on Average **(d)** Black Tiles Ratio

Fig. 6. Heatmaps of the game's board 1

significantly decreased especially when Cozmo is more assertive, according to the TKI framework. The reason for this increase in difficulty is even though the player may select the "right" action that allows Cozmo to move closer to all black tiles, it eventually will select the "down" action that will make Cozmo distance itself from the black tiles 5 and 9 and cause a drop in α_0 since the "down" action is not an action that will benefit Cozmo at some point in the round. This will also cause drops in the player's cooperation level leading to an increase of the CP_{Self} of Cozmo. The optimal trajectories for the player do not involve making Cozmo visit a single black tile. Therefore the player will, at some point in the round, frequently select actions that will not benefit Cozmo. We need also to note what happened in round 1: since the player is very competitive it is likely that α_0 had a significant drop, making it difficult for the player to have a higher score in round 2.

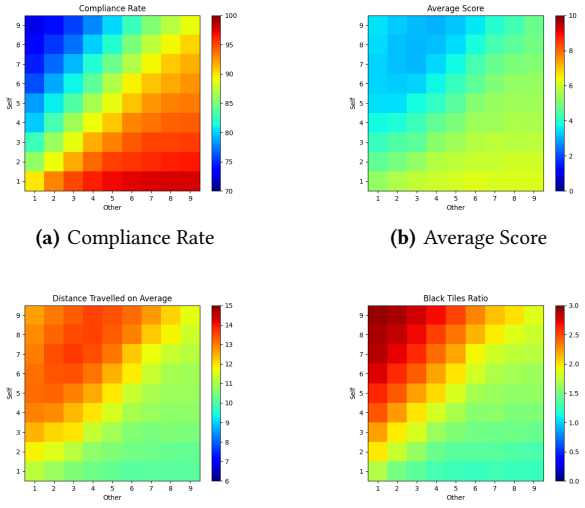
If we compare the heatmaps of board 3 (figure 8) with the ones of board 2 7 we realize that board 3 is more challenging for the player especially when Cozmo adopts a more cooperative behaviour (higher O values). This raises the question: why is board 3 more challenging than board 2? Let us take a look at heatmaps 7(d) and 8(d). For higher values O, in board 2 Cozmo typically does not need to visit any black tiles whereas in board 3 it has to visit on average between 1 and 2 black tiles, therefore making an increase on the distance travelled and lowering the score of the player. This is caused by the fact that on board 3 if the player chooses the "right" action at the beginning of the round, will cause a decrease in α_0 and in the player's cooperation level regarding Cozmo whereas at the beginning of round 2, the player will frequently select the "down" and "right" actions that benefit Cozmo, leading to an increase in α_0 and in the player's cooperation level. This means that at the beginning of a round choosing actions that benefit or not have a large influence on the unfolding of a round.



(a) Compliance Rate **(b)** Average Score

(c) Distance Travelled on Average **(d)** Black Tiles Ratio

Fig. 7. Heatmaps of the game's board 2

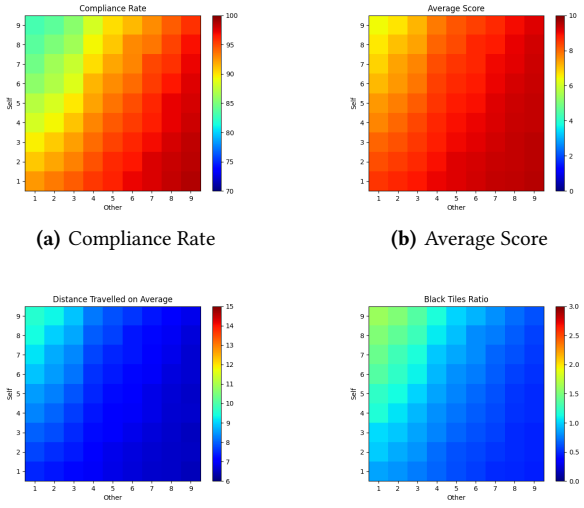


(a) Compliance Rate **(b)** Average Score

(c) Distance Travelled on Average **(d)** Black Tiles Ratio

Fig. 8. Heatmaps of the game's board 3

Finally regarding board 4, we can observe that this board is not very difficult for the player to obtain a higher score 9(b). Consequently, the compliance rate is very high for a Cozmo that is very competitive compared to boards 2 and 3 (heatmaps of figure 7(a) and figure 8(a) respectively), the distance travelled on average is low and the number of black tiles visited is also. This board is easier for the same reasons that board 1 is. Initially, the player will choose the "down" action frequently because it approximates Cozmo from the main tile and consequently, will approximate Cozmo from all black tiles and it may eventually visit the black tile 14. This raises the player's cooperation level regarding Cozmo and allows α_0 to go upwards, making Cozmo more likely to comply with the player's instructions.



(a) Compliance Rate **(b)** Average Score

(c) Distance Travelled on Average **(d)** Black Tiles Ratio

Fig. 9. Heatmaps of the game's board 4

5.4 Measures

A set of measures were used in order to evaluate the interaction between the participants and Cozmo. Some measures were collected while the participants were interacting with Cozmo (**objective measures**) and the others were obtained via a questionnaire after the interaction (**subjective measures**).

The objective measures that we are interested in are the ones that allow us to make comparisons between the events of both conditions: score, compliance rate, distance travelled and number of black tiles visited. In this experiment, the compliance rate is calculated from equation 1 where r number of times Cozmo refused to comply and n the number of instructions that the participant gave to Cozmo (the calculation of this metric does not consider the participants' withdrawal as an instruction). With these measures, we will want to compare them with the results of the simulations of this section.

$$ComplianceRate = (1 - \frac{r}{n}) \cdot 100 \quad (1)$$

To answer the second research question, we registered the cooperation level of the participants at the end of the first and fourth rounds, so that we could analyse if the participants adapted to the robot's behaviour between both conditions (participants' behaviour assessment).

The subjective measures aim to address the first research question of section 1.1 and provide insights into other relevant aspects. The questions (without considering the demographic questions), from the questionnaire, can be found in table 1. The questions are presented on a 5-point Likert scale, with a score of 1 representing

total disagreement with the statement, and a score of 5 indicating complete agreement. The final question of the questionnaire was optional and consisted of the participants writing what they would do differently if they interacted with Cozmo again in this type of scenario.

Number	Question
1	Initially, I considered the score more important to me than Cozmo's goals.
2	I wanted to gain as many points as possible.
3	I tried to cooperate with Cozmo because it refused to follow my instructions.
4	There was a lot of tension between me and the robot.
5	I feel frustrated with Cozmo because it put my score at risk.

Table 1. Questions of the questionnaire

6 RESULTS

In figure 10, we can see the average final score for each condition. A Mann-Whitney U test was performed to evaluate whether the participants' final scores differed between conditions. The results indicated that participants in the collaborative condition had a significantly higher final score than those from the competitive condition, $U = 8.000, Z = -2.543, p = 0.011$.

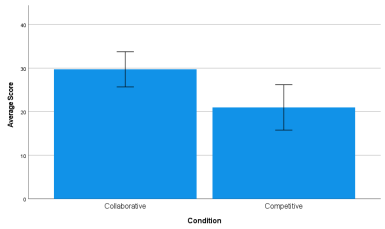


Fig. 10. Average final score

Figure 11 shows the average compliance rate considering all rounds. To assess if there were any significant differences in the compliance rate throughout all rounds for each condition, a Mann-Whitney U test was realized. The results of this test demonstrated that there were no significant differences between the compliance rate in each condition, $U = 30.000, z = -0.210, p = 0.834$.

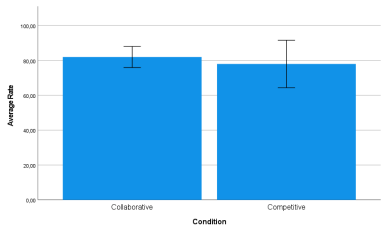


Fig. 11. Average compliance rate

Figure 12, displays the average travelled distance of Cozmo for each condition during the interaction with the participants. To observe if there were significant differences between the travelled distance in both conditions, a Mann-Whitney U test was performed. The results indicated that Cozmo travelled significantly less in the collaborative condition than in the competitive condition, $U = 7.500, Z = -2.586, p = 0.010$.

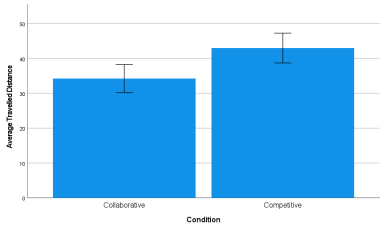


Fig. 12. Average travelled distance

In figure 13, we can see the ratio of black tiles Cozmo during the study. A Mann-Whitney U test was realized to check if the number of black tiles Cozmo visited differed by condition. The results demonstrated that Cozmo visited significantly more black tiles in the competitive condition than in the collaborative condition, $U = 7.000, z = -2.659, p = 0.008$.

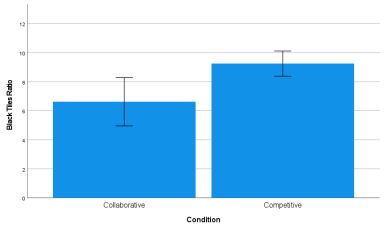


Fig. 13. Ratio of black tiles Cozmo visited

Figure 14 shows the average participants' cooperation level at the end of round 1 and at the end of round 4 for both conditions. We performed a two-way mixed ANOVA to assess participants' cooperation level towards Cozmo at the end of the first round and at the end of the final round in both conditions. The results revealed no significant differences in participants' cooperation level between the first and final rounds, irrespective of the condition they were in, $F(1, 14) = 3.429, p = 0.085$. Furthermore, there were no significant differences in participants' cooperation levels between the two conditions, $F(1, 14) = 2.614, p = 0.128$. The analysis also indicated no significant differences in how participants' cooperation level changed from the first to the fourth round when comparing the collaborative and competitive conditions, $F(1, 14) = 0.052, p = 0.823$.

Figure 15 illustrates the average scores for questions 1, 2, 3, 4[1], and 5. We conducted Mann-Whitney U tests to examine whether there were significant differences in the responses to these questions between the conditions. The Mann-Whitney U test results for question 1, $U = 27.500, Z = -0.586, p = 0.599$ indicated that

[1]Question adapted from [Amason 1996]

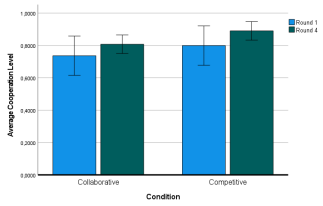


Fig. 14. Average participants' cooperation level

there were no statistically significant differences in the responses between the two conditions. The results demonstrated no statistically significant differences in the responses to question 2 between both conditions, $U = 28.500, Z = -0.413, p = 0.680$ For question 3, the results revealed no statistically significant differences in the responses between both conditions, $U = 19.500, Z = -1.367, p = 0.172$. Similarly, for question 4, the findings indicated no statistically significant differences in the responses between the two conditions, $U = 22.000, Z = -1.114, p = 0.265$. Additionally, for question 5, the results showed no statistically significant differences in the responses between both conditions, $U = 29.000, Z = -0.324, p = 0.746$.

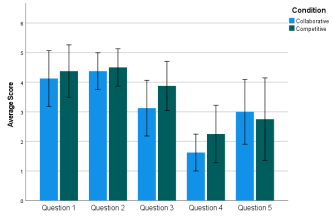


Fig. 15. Average scores for questions 1, 2, 3, 4 and 5

7 DISCUSSION

Now we will discuss the reported results answer the research questions of section 1.1. We will assume that participants exhibited a competitive behaviour since, on average, their responses to question 1 had a score above 4, signifying that they initially prioritised the score over Cozmo's goals. Furthermore, the responses to question 2 also had an average score above 4, indicating their desire to maximize their point accumulation.

Overall, there were significant differences in the final score, the distance travelled and the number of black tiles visited between both conditions of the user study. These differences were expected due to the differences in these metrics in the simulations of section 5. The only aspect that was not expected was the difficulty encountered by the participants of the competitive condition in round 1. It was also expected that there would not be significant differences in the compliance rate (except for the compliance rate of round 1 in the competitive condition): when Cozmo's aspirations, regarding the black tiles, are low Cozmo might give up on visiting black tiles, it will then try to comply with the participants' instruction leading in an increase the compliance rate at the end of the rounds. If we compare the values of the graphs with the values of the heatmaps in section 5, we can observe that both sets of values are similar except for the compliance rate, average score, distance travelled on average and the ratio of black tiles of round 1 in the competitive condition and the compliance rate of rounds 2, 3 and 4 and the black tiles ratio in the collaborative condition. These variations in the compliance rate within the collaborative condition may be attributed to the participants' ability to adjust their actions when Cozmo had previously declined to comply, a flexibility not available to the simulated player.

Research Question 1: Will people engage in an interhuman-robot conflict when they play a mixed-motive game with a robot? Although there were no significant differences in the responses to question 10 between both conditions, the average score for the competitive condition to this question is close to 4 points indicating that they tried to reach a consensus with Cozmo by trying to cooperate with it. Conversely, the responses to question 3 (with an average score of around 2 points) indicated that participants in both conditions did not experience any tension in their interactions with Cozmo. Question 5 responses have an average score of 3 around points for both conditions suggesting that that participants had mixed feelings regarding their frustration. Some may have felt frustrated to some extent, while others might not have felt that Cozmo jeopardized their scores significantly. The average score of question 4 in the competitive condition suggests that participants in this condition may have felt a sense of deprivation when it came to achieving a high score (**conflict emotionally active** as explained in 3, although the participants had mixed feelings regarding their frustration). This was evident in their attempts to cooperate with Cozmo, especially when it refused their instructions (**conflict manifestation**, as explained in section 3). Furthermore, their motivation, as indicated by their responses to question 2, was primarily focused on maximizing their points. Therefore we conclude that there were participants in the competitive condition who engaged in an interhuman-robot conflict with Cozmo. We can not conclude the same for the participants in the collaborative condition.

Research Question 2: How will people adapt to the robot's behaviour between the collaborative and competitive conditions? In our analysis of participants' cooperation levels, we found that there were no significant differences in cooperation levels between the first and fourth rounds, regardless of the participants' respective conditions. We also found no significant differences in the participants' cooperation levels between the two conditions. Participants' cooperation level remained consistent throughout the first and final rounds, and the conditions where they were in did not play a significant role in shaping cooperation levels. A noteworthy observation is that, as seen in the compliance rate analysis, participants tend to cooperate more in the first and the fourth round because they frequently choose actions that allows Cozmo to approximate the black tiles leading to an increase in the overall cooperation level. Although the cooperation level may experience a decrease when Cozmo reaches the main tile with black tiles still unvisited, it often remains relatively high. This phenomenon can be attributed to Cozmo's memory, which tends to exhibit higher cooperation levels in the latter stages of a round. As participants guide Cozmo toward the main tile, it also approaches black tiles. Notably, in the first round, Cozmo visited all black tiles while interacting with five

participants in the competitive condition. After visiting all black tiles, Cozmo's goal of reaching the main tile aligns with the participants' objective leading to an increase in the cooperation levels significantly. Additionally, it is essential to note that all participants instructed Cozmo to move to black tile 14. This last insight may suggest that participants in both conditions adapted to Cozmo's behaviour over the course of the interaction. A total of 11 responses to the last question indicate that participants, after learning how Cozmo behaves, were willing to cooperate and find a compromise between their goals and Cozmo's goals. Considering these responses and the previous insight, we can conclude that participants in both conditions adapted to Cozmo's behaviour.

8 CONCLUSION

In this paper, we focused on understanding conflicts between humans and robots namely, detecting them and what are the events before and after the emergence of a conflict. To address these issues, we designed and developed a HRI that consisted of a mixed-motive game involving a human and a robot. We applied models of previous research into the decision-making of the robot and created this interaction based on psychological research about interpersonal conflicts and research that addressed conflicts between humans and robots.

We conducted a between-groups user study with two conditions to address the previous issues. In one condition the players would play against a more competitive robot and in the other player would play against a more collaborative robot. In this user study, we were also interested in how players would adapt to the robot's behaviour in both conditions. The main findings of the study allowed us to conclude that players of both conditions adapted to the robot's behaviour by helping the robot to fulfil at least one of its goals in the latter stages of the games meaning that the players learned and adapted to the robot's behaviour. The other main finding of this study is that participants in the competitive condition engaged in conflict with the robot.

REFERENCES

Allen C Amason. 1996. Distinguishing the effects of functional and dysfunctional conflict on strategic decision making: Resolving a paradox for top management teams. *Academy of management journal* 39, 1 (1996), 123–148.

Franziska Babel, Philipp Hock, Johannes Kraus, and Martin Baumann. 2022a. Human-Robot Conflict Resolution at an Elevator-The Effect of Robot Type, Request Politeness and Modality. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 693–697.

Franziska Babel, Philipp Hock, Johannes Kraus, and Martin Baumann. 2022b. It Will not take long! Longitudinal effects of robot conflict resolution strategies on compliance, acceptance and trust. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 225–235.

Franziska Babel, Johannes M Kraus, and Martin Baumann. 2021. Development and testing of psychological conflict resolution strategies for assertive robots to resolve human–robot goal conflict. *Frontiers in Robotics and AI* 7 (2021), 591448.

Franziska Babel, Andrea Vogt, Philipp Hock, Johannes Kraus, Florian Angerer, Tina Seufert, and Martin Baumann. 2022c. Step Aside! VR-Based Evaluation of Adaptive Robot Conflict Resolution Strategies for Domestic Service Robots. *International Journal of Social Robotics* (2022), 1–22.

Henri Barki and Jon Hartwick. 2004. Conceptualizing the construct of interpersonal conflict. *International journal of conflict management* (2004).

Christoph Bartneck, Chioke Rosalia, Rutger Menges, and Inèz Deckers. 2005. Robot abuse-a limitation of the media equation. Available at: http://hdl.handle.net/10092/16925.

Andrea Bauer, Dirk Wollherr, and Martin Buss. 2008. Human–robot collaboration: a survey. *International Journal of Humanoid Robotics* 5, 01 (2008), 47–66.

Andrea M Bodtker and Jessica Katz Jameson. 2001. Emotion in conflict formation and its transformation: Application to organizational conflict management. *International journal of conflict management* (2001).

Robert Bogue. 2017. Domestic robots: Has their time finally come? *Industrial Robot: An International Journal* 44, 2 (2017), 129–136.

Sonia Brondi, Monica Pivetti, Silvia Di Battista, and Mauro Sarrica. 2021. What do we expect from robots? Social representations, attitudes and evaluations of robots in daily life. *Technology in Society* 66 (2021), 101663.

Joana Campos, Carlos Martinho, and Ana Paiva. 2013. Conflict inside out: A theoretical approach to conflict from an agent point of view. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*. 761–768.

Joana Carvalho Filipe de Campos. 2017. *Modelling Interpersonal Conflict in Multi-Agent Societies*. Ph. D. Dissertation. Instituto Superior Técnico.

Agostino De Santis, Bruno Siciliano, Alessandro De Luca, and Antonio Bicchi. 2008. An atlas of physical human–robot interaction. *Mechanism and Machine Theory* 43, 3 (2008), 253–270.

Anika Fiebich, Nhung Nguyen, and Sarah Schwarzkopf. 2015. Cooperation with robots? A two-dimensional approach. In *Collective agency and cooperation in natural and artificial systems*. Springer, 25–43.

Ron Fisher. 2000. Sources of conflict and methods of conflict resolution. *International Peace and Conflict Resolution, School of International Service, The American University* (2000).

Ronald J Fisher. 2012. *The social psychology of intergroup and international conflict resolution*. Springer Science & Business Media.

Marlena R Fraune, Steven Sherrin, Selma Šabanović, and Eliot R Smith. 2019. Is human-robot interaction more competitive between groups than between individuals?. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 104–113.

Riccardo Gervasi, Luca Mastrogiacomo, and Fiorenzo Franceschini. 2020. A conceptual framework to evaluate human-robot collaboration. *The International Journal of Advanced Manufacturing Technology* 108 (2020), 841–865.

Robert D Howe and Yoky Matsuoka. 1999. Robotics for surgery. *Annual review of biomedical engineering* 1, 1 (1999), 211–240.

Isak Karabegović. 2016. The role of industrial robots in the development of automotive industry in China. *International Journal of Engineering Works* 3, 12 (2016), 92–97.

Elena Knox and Katsumi Watanabe. 2018. AIBO robot mortuary rites in the Japanese cultural context. In *2018 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2020–2025.

Bilge Mutlu, Allison Terrell, and Chien-Ming Huang. 2013. Coordination mechanisms in human-robot collaboration. In *Proceedings of the Workshop on Collaborative Manipulation, 8th ACM/IEEE International Conference on Human-Robot Interaction*. Citeseer, 1–6.

Hiroshi Nakajima, Scott Brave, Heidy Maldonado, Masaki Arao, Yasunori Morishima, Ryota Yamada, Clifford Nass, and Shigeyasu Kawaji. 2004. Toward an actualization of social intelligence in human and robot collaborative systems. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*, Vol. 4. IEEE, 3238–3243.

Anne Reuten, Maureen Van Dam, and Marnix Naber. 2018. Pupillary responses to robotic and human emotions: the uncanny valley and media equation confirmed. *Frontiers in psychology* 9 (2018), 774.

Reinhard Selten. 1998. Aspiration adaptation theory. *Journal of mathematical psychology* 42, 2-3 (1998), 191–214.

Leila Takayama, Victoria Groom, and Clifford Nass. 2009. I'm sorry, Dave: i'm afraid i won't do that: social aspects of human-agent conflict. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2099–2108.

Kenneth W Thomas. 2008. Thomas-kilmann conflict mode. *TKI Profile and Interpretive Report* 1, 11 (2008).

Eduardo Zalama, Jaime Gómez García-Bermejo, Samuel Marcos, Salvador Domínguez, Raúl Feliz, Roberto Pinillos, and Joaquín López. 2014. Sacarino, a service robot in a hotel environment. In *ROBOT2013: First Iberian Robotics Conference: Advances in Robotics, Vol. 2*. Springer, 3–14.