



# The Influence of Interdependence on Trust Repair in Human-Agent Teams

## 1 BACKGROUND

- **Interdependence:** Relationships when multiple parties engage in joint activities
- **Trust violation:** When a teammate makes an untrustworthy action, trust violations happen
- **Trust repair strategies:** To recover from competence-based trust violation, past research has found out that expressing regret and explaining why the trust violation occurred is the most effective trust repair strategy [1].
- **Collaboration fluency:** How well a team is coordinated and how much the process is smooth and natural in a joint activity [2]

## 2 RESEARCH QUESTION

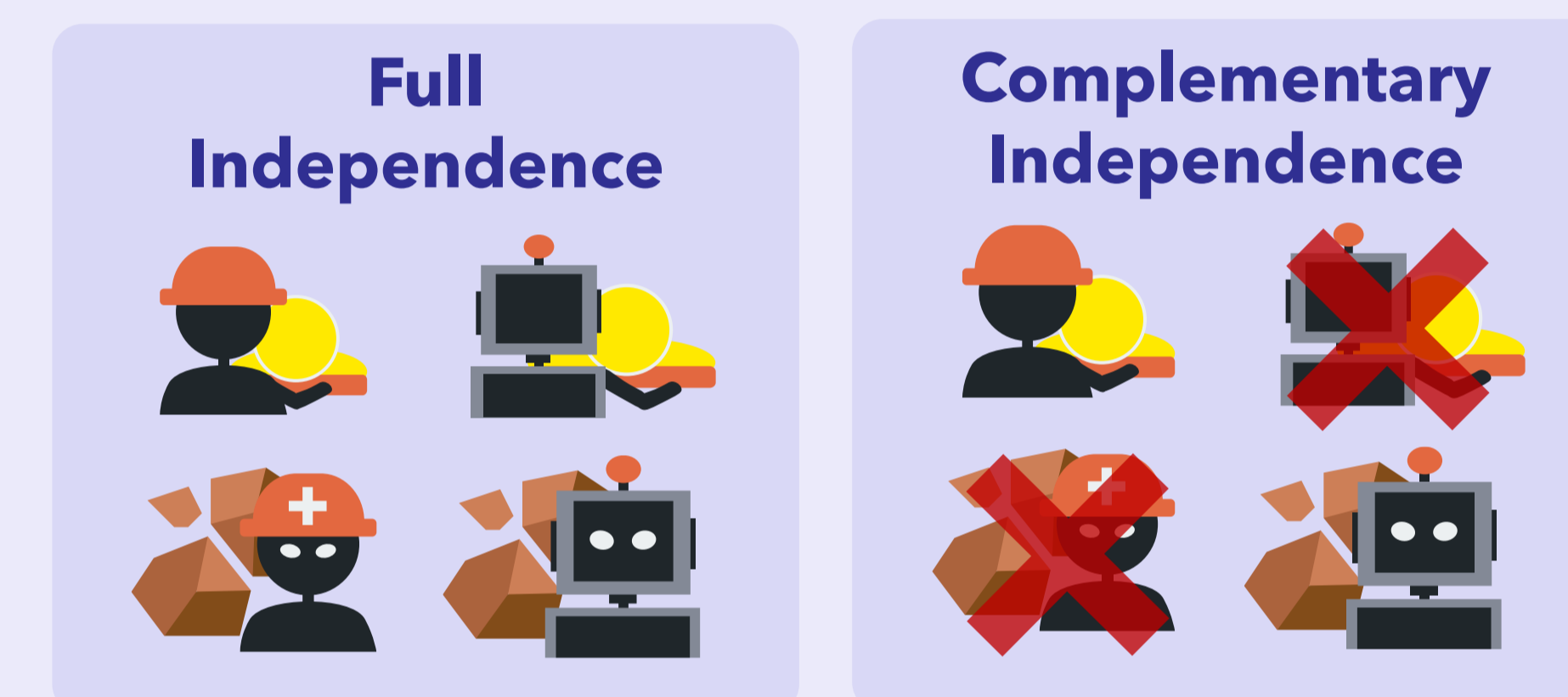
How does **full independence** and **complementary independence** in HAT influence **trust repair** and **collaboration fluency**?



Figure 1. Screenshot of the environment in god view

## 3 THE EXPERIMENT

- The human and the agent (RescueBot) is given a search and rescue mission (see Figure 1). They have to remove obstacles and rescue victims.
- RescueBot will alert the human about heavy rain in advance, but the warning may be wrong (see Figure 2). When it was wrong, RescueBot will try to recover trust with the message in Figure 3.



## 4 MEASURES

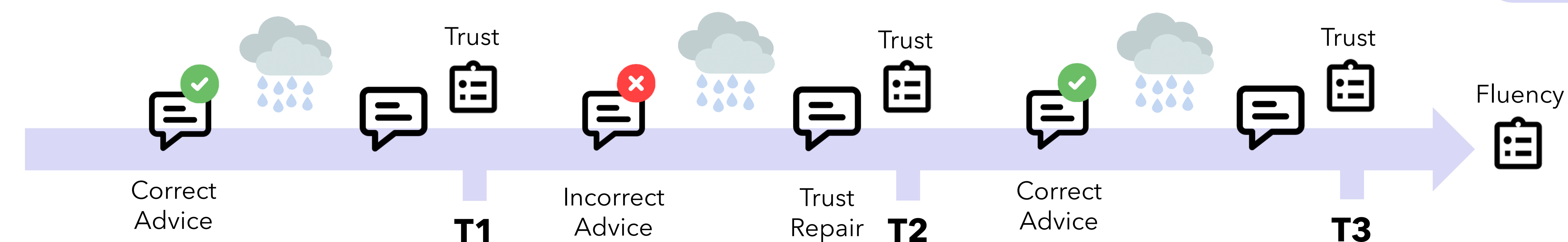
**Subjective measures** are measured through questionnaires. These are:

- Trust levels after each rain
- Collaboration fluency

**Objective measures** are logged automatically. These include:

- Score
- Completeness
- Time taken

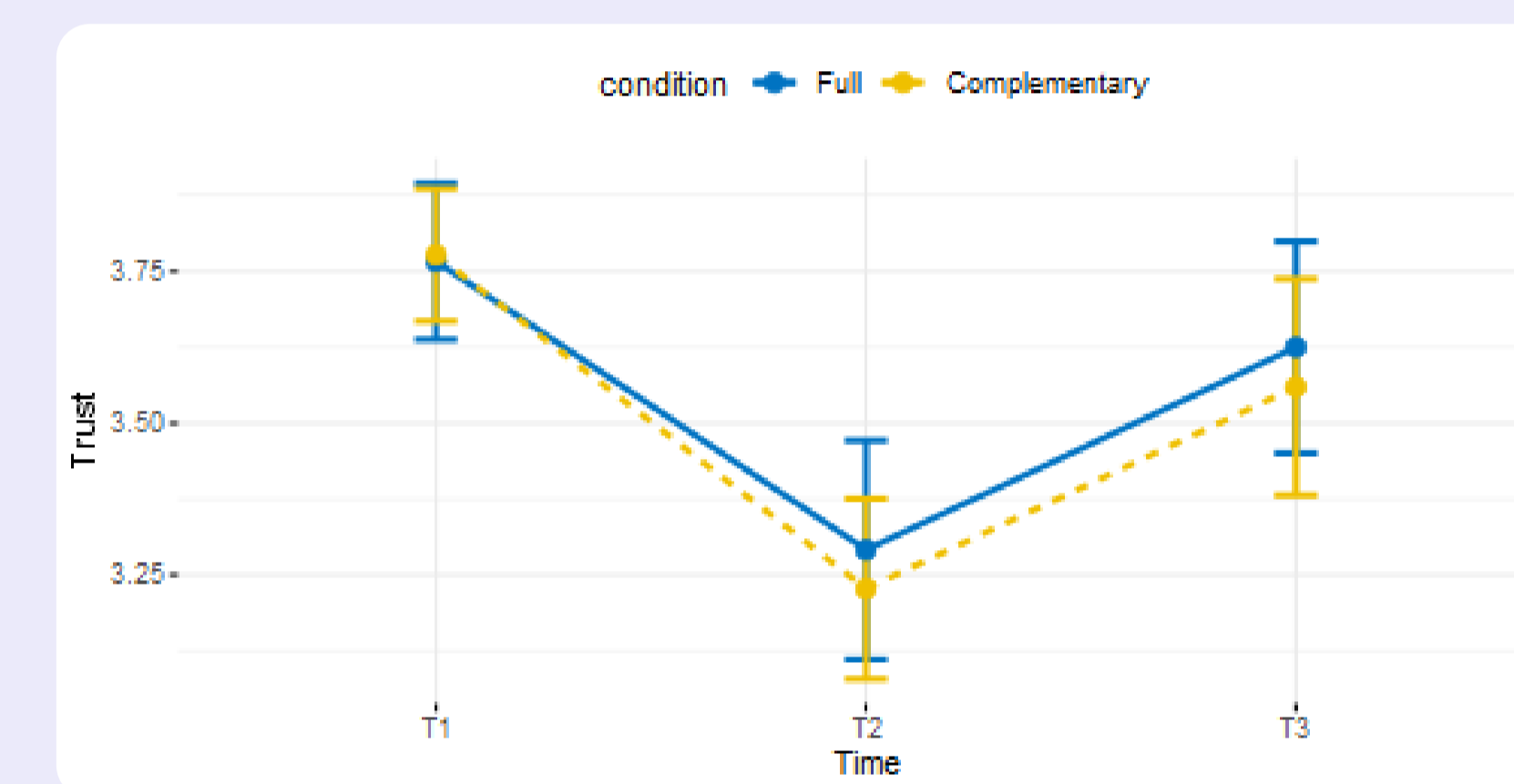
Figure 2. The timeline of the round



## 5 RESULTS

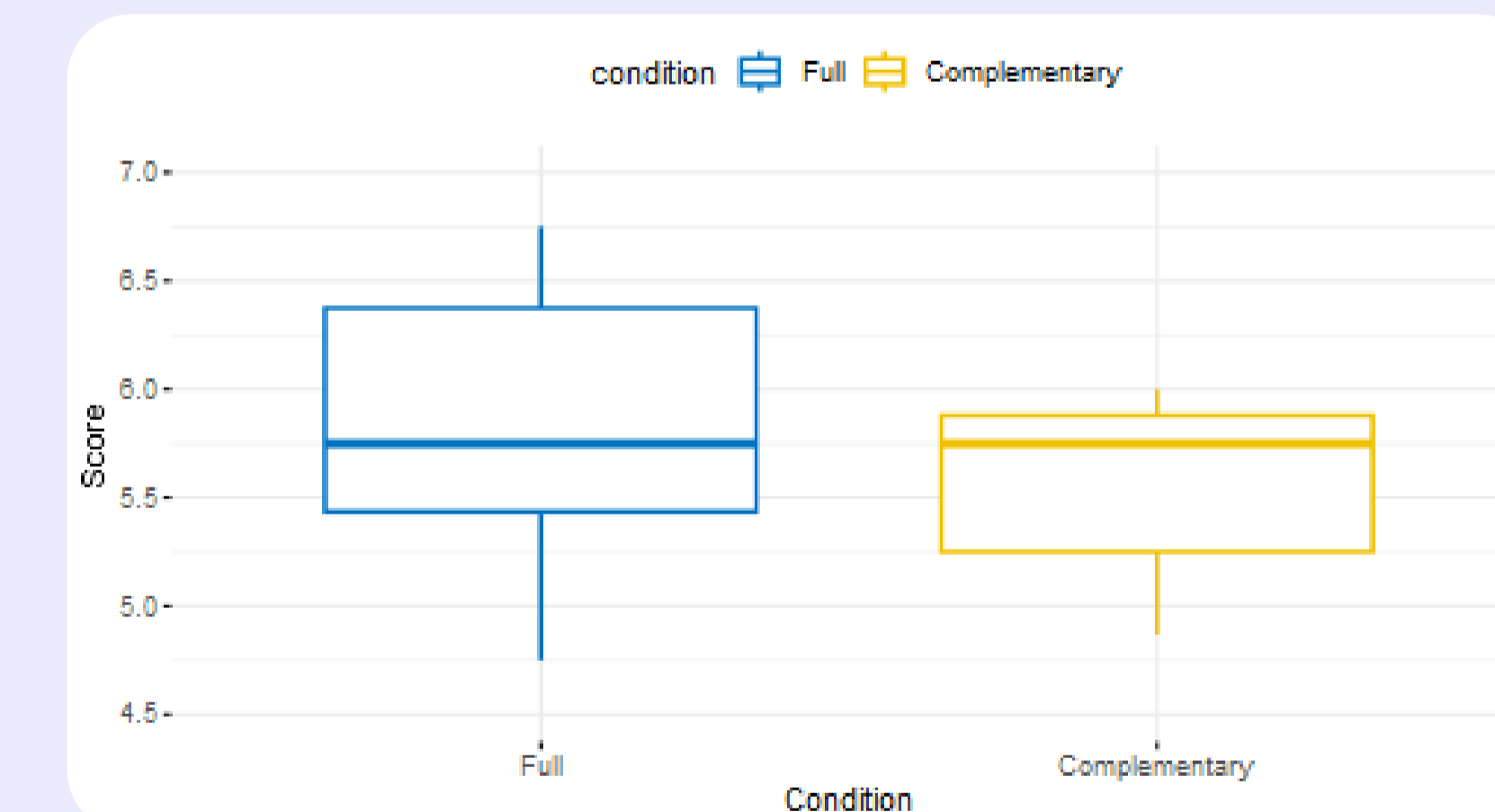
### Trust

- Two-way Mixed ANOVA was conducted
- No interaction effect between condition and time but a main effect of time ( $p = 0.005$ ).
- Significant trust violation in [T1-T2] and recovery in [T2-T3] in both conditions.



### Collaboration Fluency

- Mann-Whitney test was conducted
- No statistically significant differences ( $p = 0.307$ ).



### Performance

- Mann-Whitney test was conducted
- Significant differences found for completeness ( $p = 0.009$ ) and time ( $p < 0.001$ ) but not score

## 6 Discussion

### Trust

- The experiment did not find evidence for the relationship between interdependence and trust development.

### Collaboration Fluency

- The experiment did not reveal any correlations between interdependence and collaboration fluency.
- Although collaboration fluency and performance is not equivalent, they are closely related. However, the differences in this experiment is due to the nature of the environment setup (complementary required more waiting).

### Limitations and Future Work

- Extend spectrum of interdependence conditions
- Analyze other objective measures (e.g. advice acceptance)

[1] Matthew Johnson, Jeffrey Bradshaw, Paul J. Feltoch, Catholijn Jonker, Birna Riemdsdijk, and Maarten Sierhuis. The fundamental principle of coactive design: Interdependence must shape autonomy. In *Coordination, Organizations, Institutions, and Norms in Agent Systems VI*, volume 6541, 01 2010.

[2] Kox, E., Kerstholt, J. H., Huetting, T. A., & De Vries, P. (2021). Trust repair in human-agent teams: the effectiveness of explanations and expressing regret. *Autonomous Agents and Multi-Agent Systems*, 35(2). <https://doi.org/10.1007/s10458-021-09515-9>

[3] Hoffman, G. (2019). Evaluating Fluency in Human-Robot Collaboration. *IEEE Transactions on Human-Machine Systems*, 49(3), 209-218. <https://doi.org/10.1109/thms.2019.2904558>

My advice was wrong. The amount of rain was heavy instead of light and because of that my flood prediction was incorrect. I am really sorry.

Figure 3. The message RescueBot sends to repair trust

