Contents

1 User Study to evaluate an integrated plan and execution scheme in simulation **1**
1.1 Introduction . . . 1
1.2 Related work . . . 3
1.2.1 Questionnaires . . . 3
1.3 Study protocol . . . 3
1.4 Participants . . . 8
1.5 Study results . . . 8
1.5.1 Technical comments . . . 8
1.5.2 Statistical assumptions . . . 9
1.5.3 From execution logs . . . 10
1.5.4 From questionnaires . . . 16
1.5.5 From comments . . . 22
1.6 Discussion . . . 25
1.7 Conclusion . . . 26

I Appendix **27**

A User Study results **29**
A.1 Participants information . . . 29
A.2 Scenario Ordering per Participant . . . 29
A.3 Questionnaire answers . . . 29
A.4 Execution metrics extracted . . . 29
A.5 Participants comments . . . 30
A.6 Scenario preference . . . 30
A.7 PeRDITA questionnaire . . . 30

Bibliography **55**

Acronyms

ANOVA Analysis Of the VAriance. 9, 20

SD standard deviation. 9, 10, 11, 12, 13, 14, 18, 19

Chapter 1

User Study to evaluate an integrated plan and execution scheme in simulation

Contents

1.1 Introduction **1**
1.2 Related work **3**
1.2.1 Questionnaires 3
1.3 Study protocol **3**
1.4 Participants **8**
1.5 Study results **8**
1.5.1 Technical comments 8
1.5.2 Statistical assumptions 9
1.5.3 From execution logs 10
1.5.4 From questionnaires 16
1.5.5 From comments 22
1.6 Discussion **25**
1.7 Conclusion **26**

This chapter presents a user study validating the approach proposed in Chapter ?? using the simulator described in Chapter ??. For this purpose, several scenarios have been designed using a BlocksWorld task, and human participants were asked to collaborate with the simulated robot to evaluate its behavior. We compared our approach with a baseline behavior where the robot always imposes its decisions on the human. This study uses objective and subjective metrics to show that our approach performed significantly better than the baseline.

1.1 Introduction

To validate the approach presented in the previous chapter, we conducted a user study of more than twenty participants. The purpose of this study is two-sided. First, we want to validate our overall planning approaches. Thus, we want to show

how it allows successful collaboration with humans. Secondly, we want to validate our model of concurrent and compliant joint action, that is, showing how it allows the human to always be the leader and able to decide while the robot follows concurrently. We use a baseline where the robot imposes its decision on humans, and we show how our model allows satisfying human preferences better and is thus preferred.

We decided to conduct this study in simulation for various reasons. First, one of our assumptions is that all actions should roughly have the same duration. However, real-life robots are slow and not very reactive. Those aspects may bias the results of our study, which is focused on decision-making. Secondly, simulation allows several simplifications that are acceptable for study. Collision with the cubes has been disabled to make the robot faster both in planning and executing its arm movements. In addition, simulation allows for a perfect perception of the environment. In a real-life experiment, perception errors may occur, leading to replan and thus slower execution or even wrong decisions. Moreover, our model assumes that both agents synchronize after each step. Hence, it was easy in simulation to prevent the human from acting too soon and synchronize automatically their actions. In a real-life scenario, we could not physically prevent the participant from acting. This would imply a heavier training process for the participants to avoid desynchronizing with the robot. In practice, an additional execution supervisor should be developed to permit desynchronizing as long as they are not too big and hence, prevent the system from crashing. This would require a significant technical effort to implement.

To conduct this study, I developed a dedicated interactive simulator using a Tiago robot. In addition, the automaton described by the MoE has been implemented and integrated with the simulator to provide a proper execution and supervision scheme. Eventually, through carefully designed scenarios and using a shortened version of the PeRDITA questionnaire [Devin 2018], we gathered the feelings and impressions of the participants regarding the different robot behaviors. We also recorded logs from each executed scenario, allowing us to draw a timeline of the execution and compute objective metrics for each scenario, among which can be found the time to complete the task, the human decision time, or the time for the human to be free. Several relevant facts and conclusions can be extracted from the collected results, which are discussed in this chapter.

This chapter is organized as follows. First, the interactive simulator functionalities and operations are described. Then, the methodology of the user study is provided along with anonymous information on the participants. After that, the results obtained are presented and discussed, validating the proposed approach and our model.

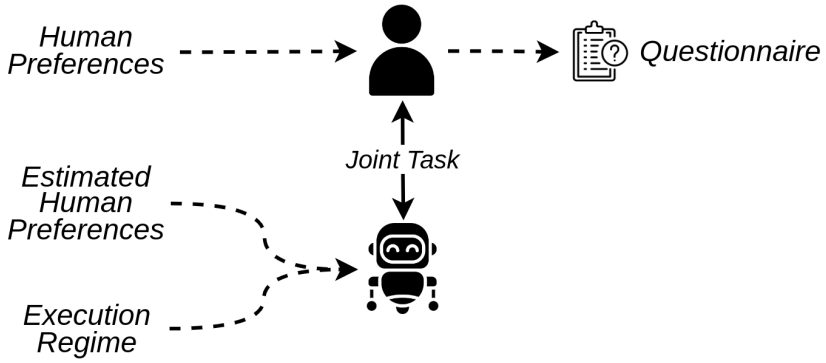

Figure 1.1: A scenario of the User Study Protocol. Each participant goes through six scenarios and answers six questionnaires to evaluate each different robot's behavior.

1.2 Related work

1.2.1 Questionnaires

Many questionnaires are used in the field of HRI. The main ones are GodSpeed, HRIES, PeRDITA, RoSAS, and Trust Perception Scale-HRI.

Each questionnaire has specificities and helps to measure certain aspects of the robot. Many include appearance items to evaluate the look of the robot. Since our focus is on robot decision-making, we decided to base our questionnaire on PeRDITA. Indeed, this questionnaire has been designed to evaluate the pertinence of robot decisions in a Human-Robot Joint Action Context, which is exactly our case. Yet, the full questionnaire is a bit heavy and also covers communication, which is not our topic in this work. That is why we decided to shorten the questionnaire by removing the section on communication and a few redundant items. Redundant items are helpful to evaluate the consistency of a questionnaire, and this has already been done in [Devin 2018]. Hence, to avoid participants getting bored and lost, we filled out the whole questionnaire. After every scenario, we kept 12 items covering the following dimensions: robot perception, interaction, collaboration, and acting.

1.3 Study protocol

In this study, each participant is made to collaborate six times with a simulated robot to achieve a shared task. Each occurrence is referred to as a scenario. The robot exhibits different behaviors in each scenario. After each scenario, the participant evaluates the robot's behavior through the PeRDITA questionnaire [Devin 2018], and logs about the execution are saved.

Beforehand, every participant answers a few demographic questions and is familiarized with the simulator functionalities through an integrated tutorial. Only

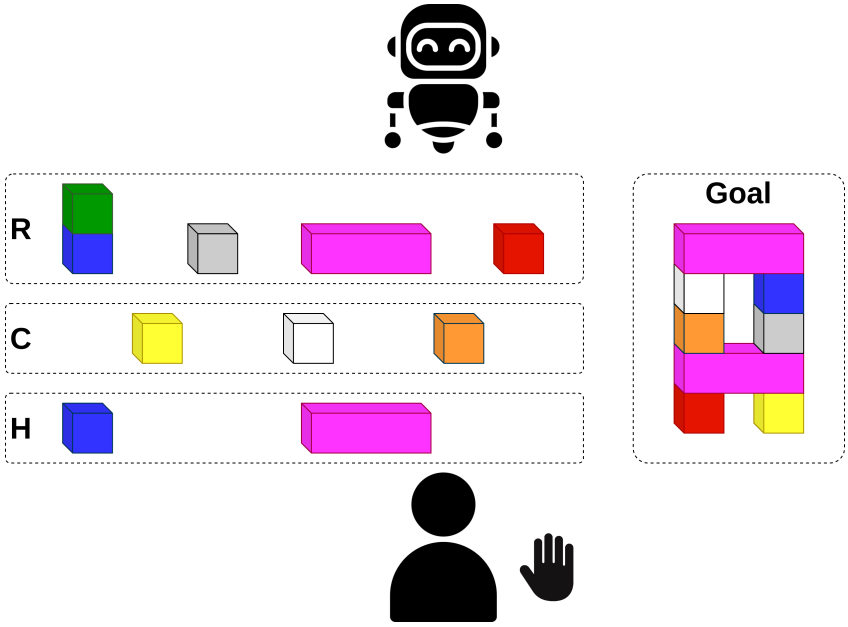

Figure 1.2: Description of the shared task to achieve in the study.

then the participants start the six consecutive collaborative scenarios, answering a questionnaire to describe the interaction every time. Eventually, every participant is asked to share their general feelings and impressions about the overall interaction with the simulated robot, and they are asked to tell which scenario they preferred the most and the least.

We now provide details about the task, the scenarios, and how the different robot behaviors are generated. The shared goal, which is stacking the cubes to match the given pattern, remains the same in all scenarios. The cube disposition on the table also does not change either. The task description is depicted in fig 1.2. For this problem, our planning approach generated a solution graph with 700 PStates/nodes leading to 6 different final states/leaves. This solution graph comprises 6839430 different possible courses of action. The length of the plans is about 19.77 ± 1.59 steps, with a minimal length of 11 steps and a maximal length of 23 steps.

To progress in the task, the agents can perform three different primitive actions, which are the following: *pick* a cube, *place* a cube in the stack, or *drop* a cube back on the table. These actions have a few preconditions, more or less intuitive, that are communicated and experienced by the participant during the integrated tutorial. First, one can *place* a cube if they hold it and if the targeted location is free and supported. That is, the cubes directly below the targeted location must be placed before being able to place a cube in the targeted location. Secondly, one can only *pick* a cube from their respective reachable zones of the table, i.e., Human and Center zones for the human and Robot and Center zones for the robot. Also, one can only pick a cube if it can be placed immediately. Thus, one cannot pick a cube "in advance" and must wait for its placement condition to be true before picking it up. For instance, both pick bars can only be picked up after the yellow and red

cubes have been placed. This rule helps to create interaction conflicts serving the purpose of this study. Moreover, although the participants found this not intuitive, they got used to it quickly, and this feeling seemed to be significantly reduced during the experiment. Third, one can *drop* a cube back on the table only if they hold it and if it cannot be placed.

For each scenario, the participant is given instructions on how to solve the task. The participants are asked to consider these instructions as their own choice and preferences regarding the task resolution and, thus, to act accordingly while collaborating. The instructions for each scenario are one of the two following. On the first hand, the participant shall act in a way to finish the task as soon as possible. Here, it consists of trying to perform as many actions in parallel as possible to progress faster. These preferences are later referred to as Task End Early (TEE). On the other hand, the participant shall act in a way to be freed as soon as possible. That is, they should finish their mandatory part of the task as soon as possible so they can leave and let the robot finish alone. Here, it consists in placing the pink bar from the Human zone as soon as possible. These preferences are later referred to as Human Free Early (HFE). On its side, the robot does not directly have access to these instructions/preferences. Hence, for each scenario, the robot is given a more or less accurate estimation of the human preferences that are communicated to the participant. Note that the participants are not aware that the robot has an estimation of their preferences, nor that this estimation can be inaccurate. This way, we created three scenarios with different pairs of human preferences and associated estimation. In the first pair, the human shall finish the task early, and the robot has a correct estimation, i.e., the robot's policy helps the human finish the collaborative task early. In the second pair, the human preferences remain the same, but the robot estimation is incorrect. The robot is trying, mistakenly, to minimize the human effort. As a consequence, the robot tends to pick cubes that the human could pick, preventing the human from acting and making the task completion longer. In the third pair, the human shall free themselves early, but the robot estimation is again erroneous. The robot will try to finish the task early while its priority is to place the first pink bar, which conflicts with the given human preferences.

Additionally, in each scenario, the robot follows one of the two following execution regimes:

- **Robot-First (RF)**: the robot always initiates actions first, and the participant takes action afterward.
- **Human-First (HF)**: the robot always lets the participant take the initiative and acts after.

The *Human-First* execution regime corresponds to the Model of Execution described in the previous chapter. At each step, the robot waits for the human's decision and will execute the best action that complies with it. The human always starts acting first, and the robot follows. On the other hand, the *Robot-First* regime

Table 1.1: Name of the six scenarios. Columns represent the preferences/estimation pairs, and the rows correspond to the execution regimes.

	Pair A TEE: correct	Pair B TEE: incorrect	Pair C HFE: incorrect
Human-First	S1	S3	S5
Robot-First	S2	S4	S6

corresponds to a naive and straightforward policy execution where, at each step, the robot directly starts executing the overall best robot action given by the policy. The robot always starts acting, forcing humans to comply. The *Robot-First* regime serves as a baseline to evaluate the proposed *Human-First* regime, described by our Model of Execution and used in policy generation. Eventually, we associate each of the three previous pairs of preferences and estimation with one of the two different execution regimes. As a result, we obtain six different scenarios with six different robot behaviors named in table 1.1.

Note that our goal is to evaluate and compare the different robot behaviors. However, at the beginning, the participants do not have any references to compare with, which can influence their answers in the very first scenarios. One solution is to ask the participants to answer all six questionnaires at the end after being familiar with the six scenarios. We consider that this option demands a too heavy mental workload to recall accurately each specific scenario and may bias the answers. As a consequence, we decided to ask the participants to answer the questionnaire after each scenario as a draft. During the experiment, they can rectify their answers to match their feelings more accurately. At the end, using the drafts, they share their final answers for each scenario. We believe this process gathers the feelings of the participants more accurately. Moreover, the ordering in which the participants encounter the scenarios is uniformly randomized to prevent any order effect.

Dimension	**Question**	**Item**
Robot perception	In your opinion, the robot is rather:	Apathetic/Responsive Incompetent/Competent Unintelligent/Intelligent
Interaction	In your opinion, the interaction with the robot was:	Negative/Positive Complicated/Simple Ambiguous/Clear
Collaboration	In your opinion, the collaboration with the robot to perform the task was:	Restrictive/Adaptive Useless/Useful Inefficient/Efficient
Acting	In your opinion, the robot choices of action were:	Inappropriate/Appropriate Annoying/Accommodating Unpredictable/Predictable

Table 1.2: PeRDITA Questionnaire: Participants have to place themselves between the two antonym items on a scale of 7.

The questionnaire filled by the participants after each scenario is a shortened

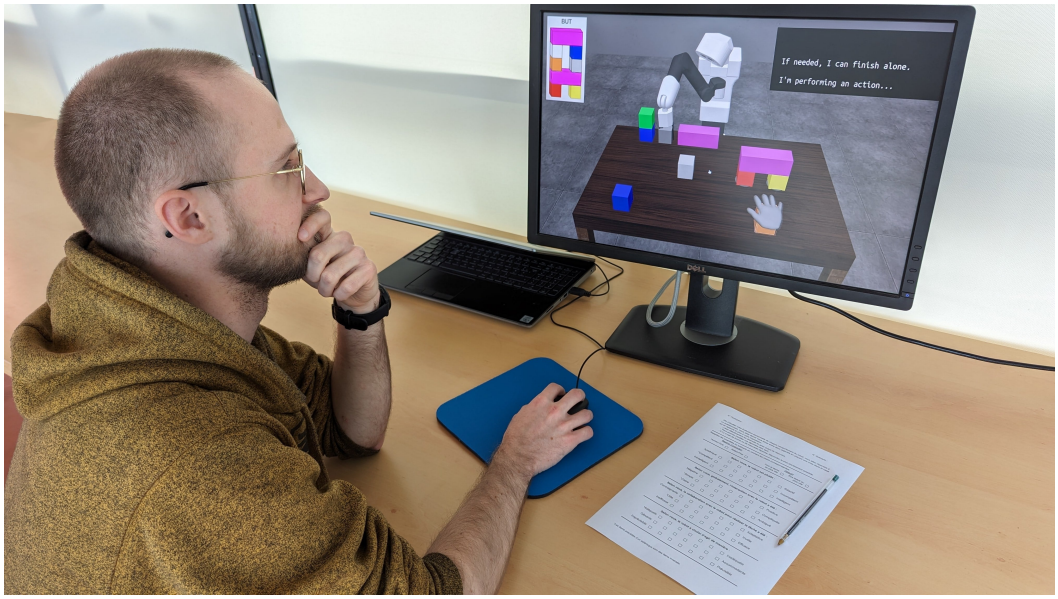

Figure 1.3: One scenario execution where a participant is collaborating with the simulated robot using the mouse. Once the task is completed, the participant is asked to fill the questionnaire on the desk with a pencil to transcribe their impressions.

version of the PeRDITA questionnaire, and its items are gathered in table 1.2. In addition to the questionnaires, for each scenario, the interactive simulator produces logs from which we extract several metrics and an overall timeline of the execution. The timeline depicts the activities and actions of each agent along the progression of the task. The subjective measures done through the questionnaire are complemented with the objective metrics extracted, such as the duration to complete the task, the number of human actions, the total duration of human inactivity, and more.

There are a few restrictions on the actions that can be performed. First, an agent can only pick cubes that can be placed immediately. This means that agents cannot pick cube in advance to anticipate each other's actions. Allowing such behavior could generate a very interesting scenario. However, here, we want to purposely generate some conflicts to evaluate the robot's behavior and reactions. Without this restriction, the agents would have too much flexibility in their actions and decisions, making it harder for conflicts to happen. Additionally, when holding a cube, the agents can only place the cube in the stack on back to its original place. As a result, the agents cannot displace the cube on the table to make them reachable to the other agent. This restriction has been added for the same reasons as the first one and simplifies the conflict generation.

The participants were collaborating with the robot using a mouse. The simulation was run on a laptop connected to a bigger screen, allowing participants to see the simulated environment clearly. After each scenario, participants answered the

printed questionnaire using a pencil. Figure 1.3 depicts the execution of a scenario where the participant collaborates with the simulated robot using the mouse. Once the task is completed and the scenario is over, the participant is asked to fill out the printed questionnaire on the desk with a pencil. For each scenario, a new paper sheet is provided to the participant.

1.4 Participants

This section shares and analyzes some information on the participants.

TODO: provide info number, age, ext, familiar with R tech, vision of robotics

1.5 Study results

In this section, we analyze the results of the study with first some technical comments regarding the experiment. After, we analyze the results obtained from the execution logs. Then, we discuss the answers to the questionnaire. Finally, we discuss the participant's comments regarding the experiment. Note that all the numeric results of the study are given in appendix A, including questionnaire answers, execution metrics, comments, and scenario preferences.

1.5.1 Technical comments

Numerous scenarios were executed in the simulator to conduct this study. More precisely, 150 scenarios were executed, and a total of 1914 steps were executed. It is interesting to share a few technical comments about how those executions.

To begin with, very few technical issues or crashes occurred during the study. About 2 or 3 crashes were due to a failure in the HMI that had happened when participants clicked at a very specific instant. I was not able to identify the origin of the issue, but this only happened a few times, considering that 1048 human actions were performed during the whole study. This means that 0.29% of the human action failed. Then, about 4 to 6 crashes occurred due to a failure of the robot arm motion controller. The arm motions were planned successfully, but the controller failed to execute the planned trajectory in the simulator, which led to a crash of the controller and the robot freezing. This kind of failure was specific to the MoveIt framework, and I could not find a solution to it. But again, those issues were quite rare considering that the robot performed a total of 1586 actions during the study. This means that 0.38% of the robot action failed. Overall, less than ten scenarios crashed during the study, i.e., less than 6% of failures. In practice, recovering from a crash was quite easy and fast. After a brief intervention of less than $30s$, the participants were able to start again from the scenario that crashed. Sometimes, this implies that the participants repeat a large part of the crashed scenario, which affects the participant's impression (less novelty effect). However, none of these crashes significantly changed the participant's actions when repeating such a scenario.

On the other hand, it is worth mentioning and discussing the durations of the different processes run by the robot. At every step, the robot has to decide which action to perform, move its head, plan its arm motion, and move its arm. First, the decision time of the robot is negligible because it is given by the policy computed previously by our planning approach. Before every step, the robot identifies the current state. Given a state, the policy dictates which action the robot should perform, including if the human action must be identified first. Since all this is pre-computed, the decision time is negligible. Head motions are also not demanding, and their execution occurs in parallel with the other robot processes. Hence, they can be neglected. However, planning the robot's arm motions is heavy computing and takes about $0.56s \pm 0.28s$. Notice that the standard deviation (SD) is quite high, about half of the average value. Indeed, the motion planning is based on algorithms using randomized exploration, making the solving time random, sometimes beginning very fast ($min \approxeq 0.001s$) and sometimes quite slow ($max = 5.37s$). Nevertheless, we were able to plan the arm motion online. Eventually, the robot arm motion durations are about $4.09s$ (max=$9.09s$, min=$1.46s$) and will be discussed more precisely below

Additionally, since the task is quite simplistic, repetitive, and deterministic, one could say that we could have pre-computed the robot arm motions in order to lighten the execution and avoid technical problems linked to motion planners. First, we insist on the fact that the arm motion failures were due to execution failure, not planning. Thus, it is unclear if pre-computed trajectories would have helped regarding those failures. Doing so would certainly make the simulator less demanding in terms of computation power. But here, we wanted to keep a generic simulator able to conduct any other task. Thus, movements could not be pre-computed.

1.5.2 Statistical assumptions

Our data are close to following a normal distribution (checked using Kolmogorov-Smirnov, Shapiro-Wilk, and Anderson-Darling tests). Thus, parametric tests can be applied, and we used both paired t-tests and Analysis Of the VAriance (ANOVA) with repeated measures to analyze the collected data, more precisely, to identify significant differences between different group of measures. In the last case, Bonferroni Post-hoc-Tests are performed to identify exactly which groups are significantly different from others.

It has been commonly assumed that a statistical test demonstrates a significant difference if a p-value lower than 0.05 is obtained. However, obtaining a value lower than 0.001 is desired. To make the p-values more legible, the following standard notation is commonly used and will be used below:

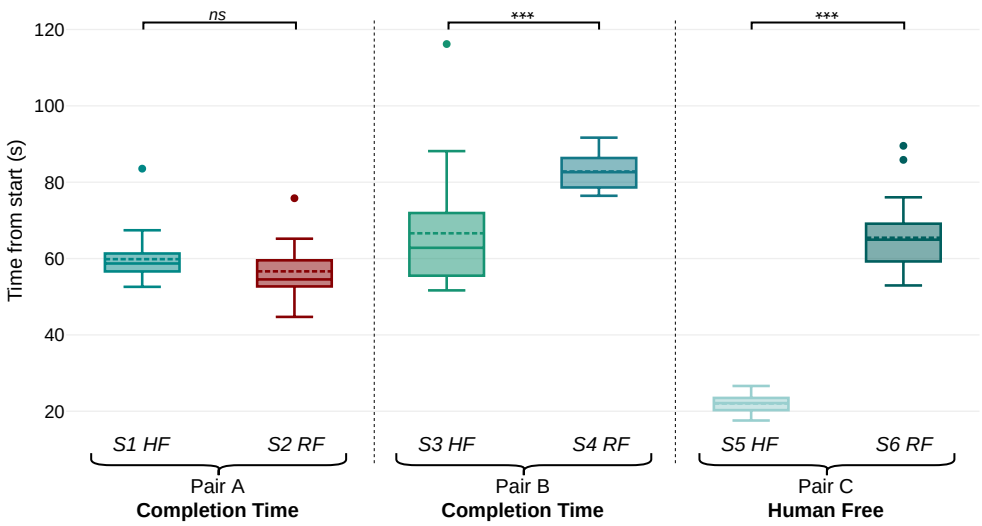

Figure 1.4: Human preference satisfaction. According to the scenarios, it corresponds either to completing the task as fast as possible (Pair A and B) or being free as soon as possible (Pair C). Using t-tests for paired samples, we can identify in pairs B and C that the criteria of preferences are significantly better satisfied. In pair A, the difference is not significant, but the completion time is slightly shorter when using RF.

$$
\begin{aligned}
p > 0.05 &\Rightarrow ns \text{ (non significant)} \\
p \leq 0.05 &\Rightarrow * \text{ (significant)} \\
p \leq 0.01 &\Rightarrow ** \text{ (very significant)} \\
p \leq 0.001 &\Rightarrow *** \text{ (highly significant)}
\end{aligned}
$$

Additionally, the value of a metric x will often be given in the following format depicting the average value M and the associated standard deviation (SD) σ: $x = M \pm \sigma$.

1.5.3 From execution logs

This section is focused on analyzing the results obtained through the execution logs saved after each scenario.

Preferences satisfaction (task completion time + time to be freed)

In this study, the human preferences consist of either finishing the collaborative task as soon as possible or being free as soon as possible while letting the robot

finish alone. Thus, to evaluate how the human preferences were satisfied, we can measure the time to complete the task in the first case and the time after which the human can leave in the second case.

Figure 1.4 depicts through box plots the corresponding relevant metric for each pair of scenarios to evaluate the human preferences' satisfaction. We used t-tests for paired samples for pairwise comparison. For each pair, the tests for normal distribution suggest that the data does not significantly deviate from normality, and thus, parametric tests such as t-tests can be conducted.

In Pair A, in addition to completing the stack, the human wants it to be completed as soon as possible. The robot has a correct estimation of human preferences. The completion time using HF and RF are shown. The completion times in S1 and S2 are roughly similar with the respective values: $59.84s \pm 5.83s$ and $56.64s \pm 6.46s$. The completion time of Scenario 1 is higher than Scenario 2. However, a t-test for paired samples showed that this difference was not statistically significant ($p = 0.055$), and there was a small effect ($d = 0.4$) according to Cohen's d [Cohen 1988] (small effect = 0.2, medium effect = 0.5, large effect = 0.8). Thus, the RF regime allowed participants to solve the task slightly faster than the HF regime. Therefore, human preferences were satisfied slightly better than using the HF regime. In both scenarios, the collaboration goes smoothly, and the task is achieved without trouble.

In Pair B, the human still wants the stack to be completed as fast as possible. However, the robot has an erroneous and adversarial estimation of their preferences. This time, the HF regime in S3 had lower values ($66.62s \pm 14.87s$) than the RF regime in S4 ($82.82s \pm 4.42s$). This difference is statistically significant ($p < 0.001$), and there was a large effect ($d = 1.07$). This indicates that in S4, the participants' preferences were significantly less satisfied than in S3. Indeed, in this pair, the robot erroneously thinks that the human wants to minimize their effort. Thus, the robot ends up trying to "*steal*" cubes from the human to prevent them from acting, thus minimizing their effort. With RF, the human has no choice and cannot act most of the time, leading to a high completion time with a low SD due to the restricted human choices, which leads to very similar executions. With HF, the robot always acts compliantly in parallel right after the human. Hence, the human is able to pick the cubes they want and that the robot wants to pick, *forcing* the robot to adapt and pick other cubes. This eventually leads to executions close to S1. However, if the human decides not to pick a cube, the robot will likely pick it, preventing the participant from acting. A few participants were distracted and let the robot pick the common cubes, leading to significantly different executions than the non-distracted participants, which explains the high SD. Comments about the feelings of the participants in each of these scenarios are given in the next subsection using the answers to the questionnaire. Overall, S4 was perceived as frustrating, and S3 was perceived similarly to S1 and S2.

In Pair C, the human prefers to be freed as soon as possible. Hence, we measured the time after which the human is not required to finish the task, i.e., the time after which the robot can finish the task alone. Scenario 5 (HF) allowed the

human to be free earlier ($22s \pm 2.35s$) than Scenario 6 (RF) ($65.45s \pm 9.08s$). This difference is statistically significant ($p < 0.001$), and there was a very large effect ($d = 4.43$). This indicates that the HF regime allowed the participants to satisfy their preferences significantly better than the RF regime. Here, the erroneous estimation of human preferences makes the robot try to place its pink bar first, which implies that the human should place their own at the top of the stack as the last cube. Such a plan forces the human to stay until the end of the task which is in direct contradiction with the actual human preferences. Hence, after placing the yellow and red cubes concurrently, both agents tend to pick their pink bar. At this point, in S5 (HF), the robot waits for the human's decision, and the human can place their bar and free themselves from the task. The robot compliantly drops its pink bar before finishing the stack alone. However, in S6 (RF), the robot does not wait for the human decision and places its pink bar before the human can do anything, forcing them to stay until the end to place the pink bar. As a result, the S6 values are significantly higher than S5. Moreover, the participants had various reactions to the frustrating robot action of placing the pink bar before them. Some remained passive until the end while holding their bar, while some others dropped it to help the robot, aiming to place the bar as fast as possible anyway. These various reactions led to various executions, explaining the high SD in S6.

Overall, RF tends to slightly better satisfy human preferences only when the estimation is correct (Pair A), yet the difference was not significant compared to the HF. On the other hand, when the estimation is erroneous, HF satisfies human preferences significantly better than RF due to how compliant the robot is when using HF. This indicates that using our model of execution instead of a simplistic baseline (RF) is beneficial for collaboration in terms of satisfying human preferences.

Ratio human optimally

The participants were given in every scenario an objective to satisfy, to consider as their own preferences regarding the task, and that should guide their behavior. However, in practice, the explicit actions to conduct were not given, and the participants were free to act as they would. Naturally, not all participants behaved in the same way. There were differences in the decision time of each, as well as in the action decisions, leading to different execution traces. Since different execution traces significantly influence the timeline metrics, it is worth discussing how the participants behaved.

First, table 1.3 depicts the number of different execution traces per scenario and overall. There were 45 different execution when considering all scenarios, which can appear quite low compared to the 6839430 possible plans. This also means that our exploration covers enough possibilities for this task. Additionally, it is worth noticing the high number of different plans in S6 and the low number in S1. In S6, the robot acts quite frustratingly, leading to various reactions from the participants. On the other hand, it seems that S1 was quite clear since participants performed only four different sequences of actions. It is also worth mentioning that since there

are only two different human objectives or preferences, we would expect only two different optimal traces, one satisfying each objective. All other 43 other obtained traces are due either to "wrong" robot decisions or suboptimal human decisions.

	Total	S1	S2	S3	S4	S5	S6
Number of different executed plan	45	4	9	10	6	7	16

Table 1.3: Number of different plans executed in each scenario and overall.

In the same manner as for the robot, an optimal human policy is generated for each scenario (considering the actual preferences given to the participant). Hence, it is possible to check at each step if the participant performed the optimal action or not and, thus, compute an optimal ratio, which is the number of optimal human actions performed divided by the total number of human actions performed. This helps us analyze the results and explain some outlier values.

Though there are no significant differences between the different scenarios, some scenarios still have a lower average optimal ratio and high SD, meaning that participants tend to have more varied behaviors in these specific scenarios. The average number of human actions per scenario is about 7, from 2 to 10.

	S1	S2	S3	S4	S5	S6
Mean	*95.52*	*95.4*	*92.04*	*98.03*	*96.74*	*87.09*
Std. Deviation	*7.56*	*7.78*	*8.28*	*4.58*	*7.65*	*13.48*
Minimum	*71.43*	*71.43*	*78.57*	*81.25*	*75*	*61.54*
Maximum	*100*	*100*	*100*	*100*	*100*	*100*

Table 1.4: Optimal human action ratio per scenario

As depicted in the table 1.4, S6 has the lowest average optimal ratio and the highest SD. In this scenario, the robot places its own pink bar even though the human holds one already, preventing the human from placing it and forcing them to drop it back on the table. This surprising behavior seems to cause frustration and confusion, which led to various human decisions and actions, and more likely to deviate from the optimal course of action. In practice, many participants get confused and are passive during several steps after the frustrating robot action. Some even remain passive almost for the whole task, waiting for the robot to stack the cubes alone until the human pink bar has to be placed. This diversity in the participants' reaction is reflected in the high SD of S6.

The low SD of S4 is also noticeable. Indeed, here, the robot tends to steal the cube from the human's reach. This behavior prevents participants from acting and, thus, from making decisions. As a result, fewer decisions are taken by the human in this scenario which results in less possible deviation from the optimal course of action.

Decision time

Participants' decision time fluctuates a lot, especially with HF. Indeed, at every step, the HF robot waits for a defined amount of time to observe the human decision and acts accordingly. Any human visual signal received interrupts this timer. This amount of time will be referred to as the HF Timeout because after it is reached, the robot considers the human to be passive. This timeout was initially set to $3s$ with the hypothesis that it should be quite small to allow fluent interaction. With a precise action in mind, humans act first and fast. Otherwise, the robot fluently takes the lead and acts first. However, during the preliminary tests, the participants felt in a rush and oppressed by this relatively low timeout. Indeed, when they did not have a precise action to perform when the step started, they did not have the time to think properly and tended to be rushed by the timer, progressing towards the timeout. Hence, we decided to increase the timeout from $3s$ to $4s$, which made it feel way more comfortable.

One could think about comparing the total (sum, cumulative) decision time over each scenario. However, since different human actions can lead to various number of steps, this is not representative.

We compare the average human decision times, measured similarly with HF and RF, and as follows. After one participant finished one scenario, we measured their decision time on each step. To do so, we first consider the time when the step begins for each step, which is signaled with text, a gaze, and a sound from the robot. Then, we consider the time when the human sends a signal by either starting an action or by waving their hand. The duration between these two times is considered as the human decision time. Note that if the human remains passive (no signal until) for a step, no decision time is computed for this specific step. Then, we extract the average decision time of the participant on the scenario from all the computed ones, compute the SD, and get the maximum and minimum values.

A one-factor analysis of variance with repeated measures showed that there was a significant difference between the variables, $F = 5.99$, $p =< .001$, with an effect size Eta squared $\eta^2 = 0.2$, which corresponds to a large effect. When doing pairwise comparisons with t-tests, the following results were obtained:

In pair A, S1 (HF) had lower values (0.66±0.41) than S2 (RF) (1.04±0.58). This difference is statistically significant ($p = 0.002$) with a medium effect ($d = 0.68$).

In pair B, S3 (HF) had lower values (0.54 ± 0.51) than S4 (RF) (0.62 ± 0.55). This difference is not statistically significant ($p = 0.551$) with a very small effect ($d = 0.12$).

In pair C, S5 (HF) had lower values (0.56±0.11) than S6 (RF) (0.91±0.46). This difference is statistically significant ($p = 0.001$) with a medium effect ($d = 0.76$).

Considering the defined scenario pairs, the decision time with RF tends to be longer. However, this difference is statically significant only for pair C (S5-S6). This is expected because when the robot places the first pink bar, the human gets confused and takes time to adapt to the situation. On the other hand, this is not reflected in S4 despite the similar confusing robot actions. Indeed, in S4, the

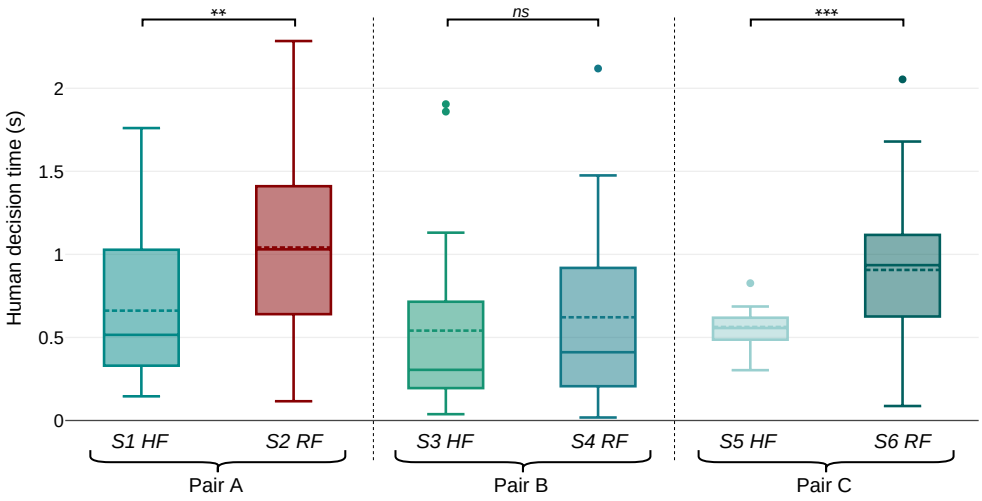

Figure 1.5: Average human decision time in the six scenarios. This decision time tends to be lower when using HF than with RF.

robot "steals" cubes from the human reach, which is confusing. Since this prevents participants from acting, no decision time can be computed.

I think the overall slower human decision time in the RF scenarios is because the human acts after the robot. This way, the human has to pay attention to the scene and the robot's action, which is longer than only looking at the scene, like in HF scenarios.

Overall, the decision time of the participants is an average of about $0.72s \pm 0.49s$.

Agent actions' duration

As depicted in fig. 1.6, the human actions are significantly faster than the robot ones on average. In addition, the duration of the robot's actions tends to fluctuate more than human ones. This can be explained by the difference in motion execution between the avatar and the robot. The human has a simplified motion planner that simply moves the hand at a constant speed and in straight lines to the cubes or the stack. However, the robot uses a real motion planner to move its arm, which is longer than human motion. The motion planning process does not always find the same solutions, nor in the same amount of time. Meaning that both the motion planning duration and the motion execution duration can fluctuate. Here, only the motion execution duration is considered in this metric. Note that to avoid having too much difference between the human and robot action durations, collisions with the dynamic objects are not considered in the robot motion planner, nor the objects' orientation. Hence, the robot can pick or place cubes from any angle and pass through the other cubes. When placing a cube, its orientation is corrected. Collisions with the table were kept, preventing the robot from picking

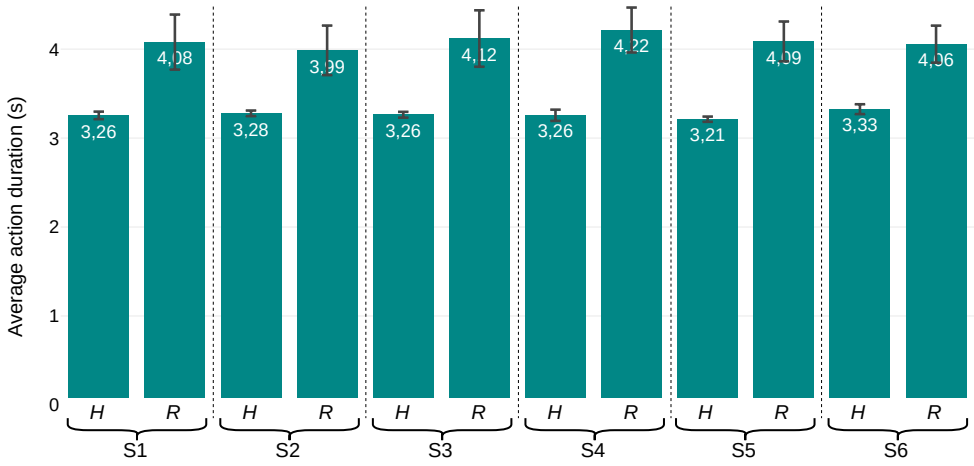

Figure 1.6: Average action duration of the human and robot agents over the six scenarios, with standard deviations. Compared to the human action durations, the robot ones tend to be longer and have various durations.

cubes from below.

Overall scenarios and steps, mean = 3.27s, the maximum human action duration is 4.63s, and the minimum is 2.55s. For the robot, mean=4.09s, the maximum action duration is 9.09, and the minimum duration is 1.46.

1.5.4 From questionnaires

This section is focused on providing the results obtained by analyzing the answers to the questionnaires filled out by the participants after each scenario.

To help the reader understand the following plots, we list here the items of the questionnaire from the table 1.6 with their associated numeric ID in table 1.5. These IDs will be used in many plots on the x-axis to analyze the questionnaire's answers.

Robot perception	**Interaction**	**Collaboration**	**Acting**
1 Responsive	4 Positive	7 Adaptive	10 Appropriate
2 Competent	5 Simple	8 Useful	11 Accommodating
3 Intelligent	6 Clear	9 Efficient	12 Predictable

Table 1.5: Questionnaire items with their associated IDs.

1.5.4.1 Overall analysis

We start by commenting on the overall questionnaire's results using relevant average values and standard deviations before having a deeper statistical analysis in the next

subsection.

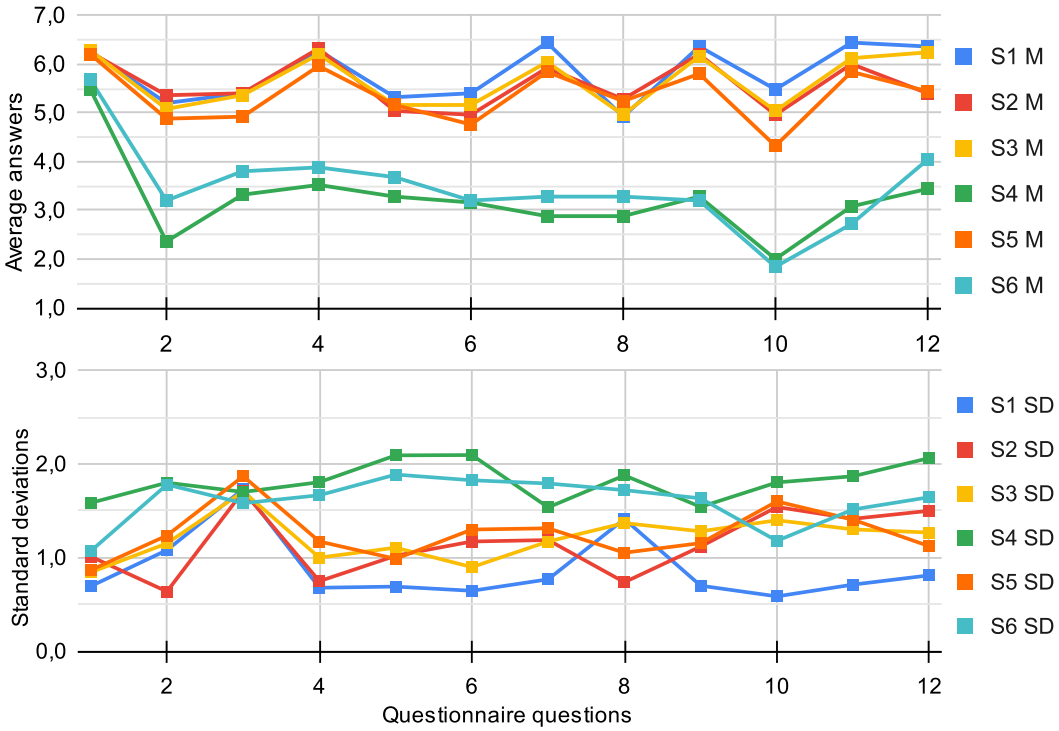

Figure 1.7: W.r.t. each scenario, average answers (M, top) and standard deviations (SD, bottom) obtained for each question of the questionnaire.

Figure 1.7 depicts the answers obtained for each question of the questionnaire w.r.t. each scenario. This figure provides a very visual overall summary of the study. On the top part, for each of the 12 questions on the x-axis, the average answers obtained are plotted for each scenario, 7 being the maximal or best value and 1 being the minimal or worst value. We can see that four scenarios obtained quite similar high answers, whereas scenarios S4 and S6 have noticeably worse answers. Those scenarios are the Robot-First scenarios of pairs B and C, where the robot has an erroneous, and even adversarial, estimation of human preferences. Note that question 1 (Q1), which evaluates the reactivity of the robot, is the only question whose answers are relatively high in every scenario. Considering the answers to all questions other than Q1, S4 and S6 seem to deviate significantly from the other scenarios, which can be analyzed as follows: First, it means that with a correct estimation, both HF and RF regimes are roughly perceived similarly. Second, an erroneous estimation does not seem to induce lower answers, and thus, despite the wrong estimation, the robot in S3 and S5 is roughly perceived similarly to the one in S1 with a correct estimation. On the other hand, when using RF, an erroneous estimation seems to have a significant detrimental impact on how the robot is perceived by the participants. All these preliminary conclusions will be confirmed in the statistical analysis below.

It is also worth commenting on the standard deviations obtained, depicted in

the bottom part of figure 1.7. The standard deviations depend a lot on the scenarios and go from 0.6 up to 2.1. There are two noticeable facts to comment on. First, question 3, evaluating how intelligent the robot is perceived, is the only question with a relatively high SD for every scenario. Participants had various definitions of "intelligence", which led to a wide range of answers. Some participants evaluated the robot's intelligence on its choices of actions, and thus, fluctuated depending on the scenario. Some others evaluated the intelligence of the robot on other criteria independent of the robot's decisions. Thus, they would rather indicate that the robot was always intelligent (or not) over all scenarios. Additionally, we can see that the SD of S4 and S6 seem to be higher than the other scenarios.

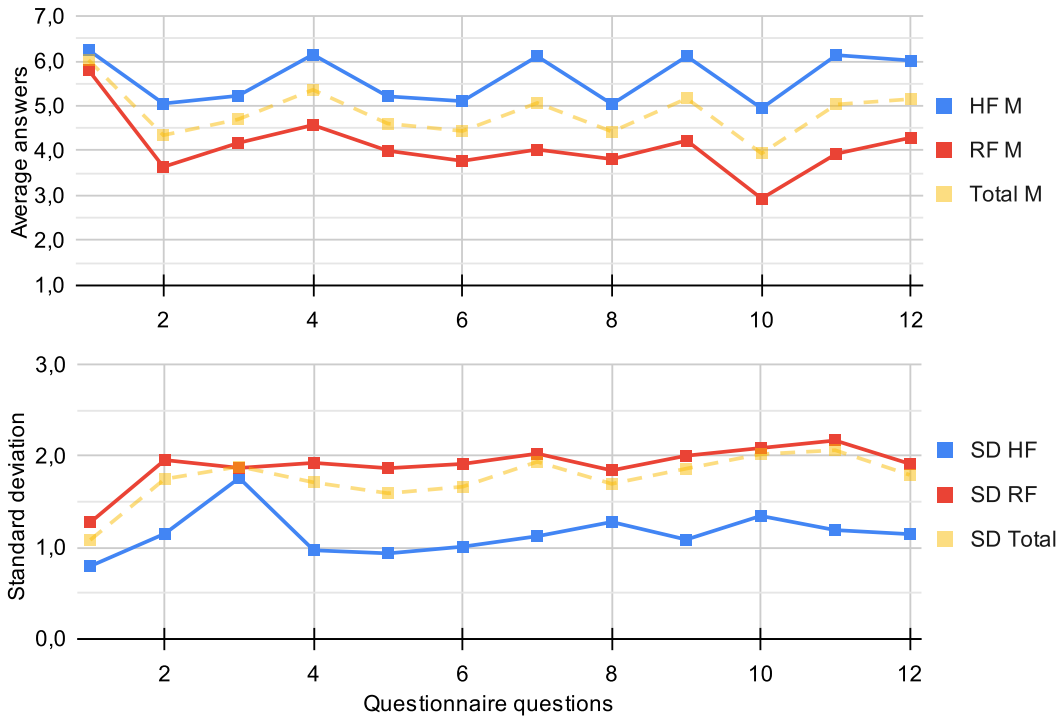

Figure 1.8: W.r.t. each regime of execution, average answers (M, top) and standard deviations (SD, bottom) obtained for each question of the questionnaire.

In contrast with the previous figure, Figure 1.8 shows the answers obtained for each question w.r.t. to each execution regime. Indeed, the HF average answers and standard deviations, shown in blue, correspond to the union of the scenario's answers using the Human-First regime, i.e., $S1 \cup S3 \cup S5$. Similarly, the RF values, shown in red, correspond to $S2 \cup S4 \cup S6$ where the Robot-First regime is used. Additionally, the combined results for all scenarios are shown in light yellow. The results shown here can be deduced from the previous figure since we already commented on all scenarios. Nevertheless, this new figure highlights more visually the difference between the HF and RF regimes.

The first noticeable fact is that when using the HF regime, the average answers to each question are better than when using the RF regime. Again, the only question

where the answers are roughly the same regardless of the regime is when questioning the reactivity of the robot. HF's average answers are relatively high for all questions which indicates that the collaborations with HF seem to be appreciated. On the other hand, RF's answers are average (around 4), indicating that collaborating with RF seems less appreciated.

Concerning the standard deviations, it is also noticeable that the answers concerning the RF regime have higher SD. This indicates that there was a wider range of answers from the participants when collaborating with the RF regime. This means that participants tend to be less certain about their answers when evaluating RF than with HF. This can be expected because when facing the HF regime, the collaboration is quite positive overall. Therefore, participants concentrated their answers on the higher part of the scales. However, the frustrating robot actions due to the RF regime degraded the collaboration, and participants had to evaluate how bad this degradation was. Some participants were more emotionally affected than others by the robot's actions. Hence, it led to a wide range of answers. Another interesting fact is that regardless of the regime, question 3 has a high SD. This question evaluates the *intelligence* of the robot. Participants had various definitions of intelligence, which is reflected in this high SD. Indeed, some participants evaluated the robot as unintelligent because of its nature and, thus, regardless of its actions or the scenario. Others perceived less intelligence when the robot performed frustrating actions. Also, we can see that the SD regarding the reactivity of the robot is quite close and low for both regimes. This is a consequence of the very high average values regarding Q1 for both regimes.

Eventually, I would like to insist on the average answers concerning the RF regime. Even if these answers are mediocre, it is important to note that they are not very low. In comparison, consider another robot behaving completely erratically. This robot would randomly pick and place cubes around itself. Consequently, the robot would neither help humans nor solve the task. On the contrary, it might only disturb the human trying to achieve the task. The robot could make the task impossible for the human by picking a relevant cube and never placing it or removing already well-placed cubes from the stack. Here, during one step, the RF regime forces humans to comply with the robot's decisions, which can be frustrating. However, the robot takes human action into account and adapts its actions accordingly for the next step. This is thanks to our planning approach, which is common to both regimes and explores every possible human action to generate the robot policy. Thus, we can say that our planning approach seems to benefit the collaboration and interaction between the human and the robot.

1.5.4.2 ANOVA Analysis

So far, we only conducted a preliminary analysis of the questionnaire's answers using only the average values and standard deviations. Now, a statistical analysis must be conducted to confirm the preliminary comments of the previous subsection. For each question, and thus each item of the questionnaire, we performed

an analysis of the variance with repeated measures. Each analysis led to p-values $<= 0.001$, indicating that there is a significant difference between the six scenarios. To evaluate the strength of this significant difference, the effect size Eta squared η^2 has been calculated where the limits are .01 (small effect), .06 (medium effect), and .14 (large effect). However, ANOVA tests can only indicate if there is a significant difference between N-samples, but it is only of interest to identify between which exact group that difference exists. In the Bonferroni post-hoc test in an ANOVA with repeated measures, multiple t-tests are calculated for dependent samples. However, the problem with multiple testing is that the so-called alpha error (the false rejection of the null hypothesis) increases with the number of tests. To counter this, the Bonferroni post-hoc test calculates the obtained p-values times the number of tests. The obtained p-values indicate in a pairwise manner between which samples the significant difference exists. The results from the ANOVA and Bonferroni post-hoc test are shown in table 1.6.

As suggested by the preliminary analysis, the robot's reactivity is the item with the lowest difference. The ANOVA indicates that answers regarding the reactivity are significantly different with a large effect over the six scenarios. However, compared to other items, the effect size η^2 is quite low, indicating that this difference is less significant than for other items. Also, the Bonferroni post-hoc test was not able to identify where exactly the difference exists, which means again that this difference is not very significant in the end.

Besides reactivity, all other items have significant differences according to the scenario, which can be exploited using the Bonferroni post-hoc test. Indeed, the pairwise comparisons indicate the existence of major significant differences for every question in the following pairs: S1-S4, S1-S6, S2-S4, S2-S6, S3-S4, S3-S6, S4-S5, and S5-S6 (in bold in the table). Having in mind that S4 and S6 are the two scenarios using Robot-First with erroneous estimations, we can see that erroneous estimation with RF systematically leads to significant differences compared to all other scenarios using HF or RF with a correct estimation (S4-S1, S4-S2, S4-S3, S4-S5, and S6-S1, S6-S2, S6-S3, S6-S5). This is a clear indicator that the RF regime is very sensitive to the estimation of human preferences. Thus, an erroneous estimation is significantly detrimental to the collaboration and the overall interaction.

Only a few other significant differences exist in S1-S5 and S3-S5. Indeed, in S5, the robot's actions were perceived as more or less significantly less appropriate ($p = 0.004$) and predictable ($p = 0.004$) than in S1, slightly less predictable ($p = 0.017$) than in S3. Indeed, in S5, the HF robot surprisingly picks up its pink bar while the human picks up its own whereas the human wants to place their bar to be freed from the task. Using HF allows the human to place their bar anyway, but the robot actions were therefore perceived as less predictable and appropriate than in S1 or S3. We can also notice that despite S3 having an erroneous estimation in contrast to S1, there is no significant difference in answers for each question between S1 and S3. This indicates that using the HF regime allows the robot to be way more robust to erroneous estimation than RF. However, the few existing significant differences among the HF scenarios indicate that an erroneous estimation is still noticeable and

	ANOVA		Bonferroni Post hoc tests													
	p	η^2	1-2	1-3	**1-4**	1-5	**1-6**	2-3	**2-4**	2-5	**2-6**	**3-4**	3-5	**3-6**	**4-5**	4-6	**5-6**
Responsive	***	*0.16*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*
Competent	***	*0.49*	*ns*	*ns*	***	*ns*	***	*ns*	***	*ns*	***	***	*ns*	**	***	*ns*	*
Intelligent	***	*0.33*	*ns*	*ns*	**	*ns*	*	*ns*	**	*ns*	**	**	*ns*	***	*	*ns*	*
Positive	***	***0.65***	*ns*	*ns*	***	*ns*	***	*ns*	***	*ns*	***	***	*ns*	***	***	*ns*	***
Simple	***	*0.34*	*ns*	*ns*	**	*ns*	**	*ns*	**	*ns*	*ns*	*	*ns*	*	**	*ns*	*
Clear	***	*0.41*	*ns*	*ns*	***	*ns*	***	*ns*	**	*ns*	***	***	*ns*	***	*	*ns*	**
Adaptive	***	***0.62***	*ns*	*ns*	***	*ns*	***	*ns*	***	*ns*	***	***	*ns*	***	***	*ns*	***
Useful	***	*0.43*	*ns*	*ns*	***	*ns*	*	*ns*	***	*ns*	***	***	*ns*	**	***	*ns*	***
Efficient	***	***0.62***	*ns*	*ns*	***	*ns*	***	*ns*	***	*ns*	***	***	*ns*	***	***	*ns*	***
Appropriate	***	***0.62***	*ns*	*ns*	***	**	***	*ns*	***	*ns*	***	***	*ns*	***	***	*ns*	***
Accommodating	***	***0.68***	*ns*	*ns*	***	*ns*	***	*ns*	***	*ns*	***	***	*ns*	***	***	*ns*	***
Predictable	***	*0.45*	*ns*	*ns*	***	**	***	*ns*	*	*ns*	*	***	*	***	**	*ns*	**

Table 1.6: Significant differences in the questionnaire answers between the different scenarios. For each item of the questionnaire are shown the overall p-value and η^2 (effect size) obtained after an ANOVA. Additionally, the p values obtained after conducting Bonferroni Post-hoc-Tests are shown to identify in a pair-wise manner which scenarios were significantly different from others. As depicted, scenarios S4 and S6 are distinguishable from the others, and their evaluation is significantly different on all the measured aspects (expect reactivity).

can still have a detrimental influence. Thus, the robot cannot fully rely on being reactive and compliant to human actions. Estimating human preferences accurately to plan the robot's actions appropriately is mandatory for optimal collaboration.

Looking at the eta squared η^2 of every question, one can notice that we can group the items into four groups.

1. $\eta^2 = 0.16$: First, the Reactivity item is alone with the lowest effect size. This group does not help to distinguish the regimes.

2. $\eta^2 \simeq 0.33$: Secondly, we can state that when using RF, the robot was perceived as slightly less intelligent, and the interaction as slightly less simple than when using HF.

3. $\eta^2 \simeq 0.45$: Then, due to the moderate effect size, we can state that when using RF, the robot was perceived as moderately less competent, the interaction as moderately less clear, the collaboration as moderately less useful, and the robot actions as moderately less predictable.

4. $\eta^2 \simeq 0.64$: Eventually, with a high effect size, when using the RF regime, the interaction was perceived as significantly less positive, the collaboration as significantly less adaptive and efficient, and the robot actions as significantly less appropriate and less accommodation. Since the five items from the last group are the ones with the highest effect size, they can be seen as the main characteristics differentiating the HF from the RF regimes and thus are highlighted in the table.

1.5.5 From comments

At the end of the experiment, every participant was asked two questions, gathering their general impressions. First, they were asked to comment on the overall experiment and robot interaction they just had. Secondly, participants were asked to indicate which scenario they preferred the most and the least.

Participants' comments concern several aspects of the experiment and are worth discussing. Overall, the comments confirm the outcome of the statistical analysis, and they also provide some feedback about the overall experiment protocol and conditions, especially regarding the simulator itself. The comments are discussed per categories below.

1.5.5.1 Simulation

First, most of the participants found the experiment and the simulation to be a good experience. They felt committed and active during the different scenarios. The simulation has been described several times as clear, simple, pleasant, intuitive, captivating, and funny. Some participants mentioned that it was like a video game and enjoyed it. These comments suggest that collaborating with a simulated robot has been appreciated, and they raise the question of whether the participant had felt the same way in an experiment with a real robot.

1.5.5.2 Display

Some participants think that the simulation display had too much information (goal + scene + text prompt), and a few had trouble reading the text prompt. Yet, the robot can be seen well. Indeed, the text prompts could be a bit fast and written white on black, which can be disturbing when not used to. However, I believe this did not affect much the executions, maybe slightly the decision times.

1.5.5.3 General

A few participants would have appreciated the robot giving instructions and guidance regarding the actions to perform. This is linked to another comment saying that using the Robot-Fist regime with correct estimation felt better because it makes the task simpler. If the human trusts the robot, it can be appreciated to let the robot compute the optimal plan and just follow the robot's instructions, lowering the cognitive load of the human. Indeed, some participants consider that they made mistakes and that they could have acted better in some scenarios.

1.5.5.4 Steps

The step synchronization was not appreciated by everyone. Some participants found this kind of synchronization useful as it structured the collaboration. However, many others found this a bit confusing at first and frustrating because they had to wait for the robot's actions to be done before being able to act again.

1.5.5.5 Task

The task was found to be clear and quite simple. One participant said that they felt significant emotions such as satisfaction and frustration and that if the task was less abstract and more real, these emotions would have been enhanced. Moreover, another participant said that in such simple tasks, humans think they know better how to solve the task than the robot. Thus, the robot should follow human decisions in such cases, with a hierarchy relation. A few participants also mentioned that performing the last action, i.e., placing the last cube, is very satisfying. These people liked the robot adapting to allow them to do so. The fact that both agents must perform actions to solve the task makes the collaboration relevant and useful.

1.5.5.6 Action

Many participants said that not being able to pick cubes in advance is not natural and, at first, it is confusing, frustrating, and a bit complicated. Yet, they also said that they got used to it quite fast. One also said that they felt obliged to act at every step. Indeed, a majority of the participants were signaling their passivity to the robot even when they were not able to act. About the movements, one participant stated that the actions were stiff and rigid, in contrast to being able to drag and drop the cubes thanks to the physics simulation. Additionally, the lack

of collision with the cubes felt a bit unrealistic but not very confusing. On the other hand, another participant said that the robot's movements seem real. This is probably because we used an online motion planner to move the robot arm. Thus, the movements were not always optimal.

1.5.5.7 Objective

A few participants said that the objective of "trying to be free early" is a bit frustrating since they would like to keep acting, even if not necessary. It was hard for them to consider this objective as their personal preference and, thus, to act accordingly. Additionally, one participant mentioned that Scenario 5 creates double satisfaction: being free early (preferences) and fulfilling the task.

1.5.5.8 Regimes

One participant said that they did not see much difference between the two execution regimes, HF and RF. The same participant could not indicate which scenario they preferred at the end. Moreover, some participants also indicated that the difference between HF and RF was unclear at first. However, once used to the task and the scenarios, the difference becomes clearer, and before the end of the experiment, most of the participants had a clear idea of each regime and even apprehended the RF one.

1.5.5.9 Being in control

Participants indicated that when using HF, they felt in control and free to decide which action they performed and that the robot was adapting to their decisions and actions, which was appreciated. In contrast, when using RF, participants did not feel in control and were forced to adapt to the robot's decisions. Even when the robot's decisions are good, the lack of control is uncomfortable. One participant stated that they disliked when the robot took the initiative because the robot could be wrong.

1.5.5.10 Human-First (HF) regime

Most of the participants enjoyed the HF regime and stated that they were able to fulfill their objective with it. Some comments qualify the HF regime as slower than RF and sometimes inconsistent. The latter is mostly referring to the robot picking up the pink bar in S5. However, especially when used to it, HF has been qualified as smooth, efficient, interesting, predictable, and less frustrating than RF. Several participants mentioned that they enjoyed being able to predict the robot's behavior, proving that having predictable behavior is crucial for a seamless collaboration. It has been mentioned that HF makes less wrong choices than RF. Moreover, some participants said that compared to the RF regime with a correct estimation, HF is less efficient. However, in pairs B and C, HF is more efficient than RF.

1.5.5.11 Robot-First (RF) regime

RF, bad, not in control, bad choices: having to drop bar + using common resources first

Every participant had an overall negative opinion regarding the RF regime. The latter has been qualified as very frustrating, confusing, constraining, unpredictable, inefficient, and even adversarial. A significant number of participants stated that finishing the task quickly with RF could be great, fast, efficient, and less cognitively demanding despite the lack of control. Also, a few participants noticed that even if, during a specific step, the human is forced to comply with the robot's actions, the robot takes into account the human action and adapts its behavior in the next step. However, it has been said that RF does not consider the human's objective or preferences. Participants really disliked when the robot forced them to drop a cube back on the table (pink bar in S6) and when the robot picked cubes in the middle zone instead of its own zone. The latter was perceived as the robot stealing the cubes from the human. Due to those frustrating robot actions, the RF regime was putting the participants in an adversarial setup, and the robots were explicitly qualified as "enemies". Some participants said that they were more focused on preventing the robot's mistakes than on the actual task.

1.6 Discussion

There are several elements to discuss in this study, including several participants' comments.

First, we decided to conduct the simulation study instead of implementing the system on a real robot. The main reason for this choice is that using a real robot would have required significant additional work. We could not physically force the human to synchronize with the robot according to our step-based model. Therefore, many executions could have failed, and thus, the study would have been longer to conduct. Real robot motions are likely to be slower than our simulated ones, which could also result in bias. An in-between solution could be to use Virtual Reality (VR) to make the participant more immersed.

Our study focused on the robot decisional aspect of collaboration. For this reason, we simplified a few elements of the simulation, such as collisions and physics. Despite being noticeably less realistic than real life, participants seem to find it adequate.

The text prompts raised several questions about which information to show, when, and how. A few participants stated that it was sometimes hard to read the prompts. Further study on these prompts would be necessary.

Due to our limited number of participants, we had to focus on a single collaborative task in order to obtain significant results. Asking participants to solve various tasks would have been too cognitively demanding. Therefore, additional results from another collaborative task would strengthen our results.

Finally, using another baseline would also benefit our study. We mentioned a

possible baseline where the robot behaves completely erratically. Such a baseline would have highlighted that the RF regime is not so bad and permits always solving the task, and in an efficient way when human preferences are correctly estimated.

1.7 Conclusion

Thanks to this study, we aimed to validate the overall planning approach and the model of execution Human-First, which is critical to our approach.

After statistically analyzing the execution log data as objective metrics and the questionnaire answers and participants' comments as subjective metrics, we can confidently state that this study successfully validates both our planning approach and our model of execution.

Indeed, we have solid proof that the HF regime gives humans control over the execution, which was significantly appreciated. The participants perceived the robot as accommodating, adaptive, and acting appropriately while being predictable. HF also helps to satisfy better human inner preferences, which makes it more robust to erroneous estimations and thus more enjoyable. On the other hand, we show how the RF regime can be greatly appreciated when estimating human preferences correctly. However, we demonstrate how erroneous estimations strongly harm collaboration and interaction using the RF regime. Hence, the Human-First regime is preferred and allows for achieving smooth, efficient, and positive collaborations. Nevertheless, thanks to our planning approach, we also show that the RF regime always solves the task with humans, and thus, it is always helpful. Additionally, it also always adapts to human action in the next step.

Part I

Appendix

Appendix A

User Study results

This appendix provides the raw results obtained through the User Study described and discussed in chapter 1.

A.1 Participants information

Anonymized information on the participants is shared in the tables A.1 and A.2. It includes the date of each participation, the age and gender of the participant, their opinion about robotics (1=negative, 5=positive), if they are from my laboratory (LAAS = yes, EXT = no), if they are familiar with robotics and task planning, and if they already interacted with a robot, if so which ones. The comments are given in the participants' language, hence, mostly in French, except for participants 10 and 16.

A.2 Scenario Ordering per Participant

The ordering in which each participant encountered each scenario is shown in table A.3. These orderings have been randomized to avoid order effect and follow a uniformed distribution.

A.3 Questionnaire answers

The participants' answers are provided in the 3 tables A.4, A.5, A.6. For each participant, the answer on a Likert scale from 1 to 7. Be careful, by default, 1 corresponds to the worst answer and 7 to the best. However, some questions have inverted scales in the questionnaire. The inverted questions are : Q2, Q5, Q6, Q8, and Q9. In the rest of the manuscript, especially in chapter 1, the inverted scales are inverted to have a uniform and more legible representation.

A.4 Execution metrics extracted

The extracted execution metrics are abbreviated corresponding to the table A.7. The metrics values are shown in ten tables from A.8 to A.17.

A.5 Participants comments

The participants' comments about the experiment are shown in the tables A.18, A.19, and A.20. Since these comments were given verbally they are here described as note-taking, in French which is the language spoken by a large majority of participants, and using as much as possible the participant's own words.

A.6 Scenario preference

The results obtained after asking participants which scenario they preferred the most and the least are shown in the table A.21.

A.7 PeRDITA questionnaire

The PeRDITA questionnaire filled by the participant after each scenario can be found in the next two following pages. Both the original French version and an English translation done by myself and based on other already translated versions.

PeRDITA Questionnaire FR

N° Participant : N° Scénario :

Afin d'étudier votre évaluation personnelle de chaque comportement du robot, vous allez répondre à un questionnaire. Vous allez devoir vous situer entre deux adjectifs en plaçant une croix dans la case qui se rapporte le plus à votre impression.

- Lisez attentivement les énoncés en gras avant de répondre.
- Vous pourrez modifier vos réponses plus tard.
- Il n'y a pas de bonne ou de mauvaise réponse. Répondez le plus sincèrement possible.

Rappelez les propriétés du scénario en cochant les cases correspondantes:

Régime d'exécution

Human-First ☐ Robot-First ☐

Objectif

Finir la tâche au plus vite ☐ Être libéré au plus vite ☐

Selon vous, le robot est plutôt :

Apathique	☐	☐	☐	☐	☐	☐	☐	Réactif
*Compétent	☐	☐	☐	☐	☐	☐	☐	Incompétent
Inintelligent	☐	☐	☐	☐	☐	☐	☐	Intelligent

Selon vous, globalement, l'interaction avec le robot a été :

Négative	☐	☐	☐	☐	☐	☐	☐	Positive
*Simple	☐	☐	☐	☐	☐	☐	☐	Compliquée
*Claire	☐	☐	☐	☐	☐	☐	☐	Ambiguë

Selon vous, la collaboration avec le robot pour réaliser la tâche a été :

Contraignante	☐	☐	☐	☐	☐	☐	☐	Adaptative
*Utile	☐	☐	☐	☐	☐	☐	☐	Inutile
Inefficace	☐	☐	☐	☐	☐	☐	☐	Efficace

Selon vous, le robot a choisi d'agir de manière :

*Adéquate	☐	☐	☐	☐	☐	☐	☐	Inadéquate
Gênante	☐	☐	☐	☐	☐	☐	☐	Accommodante
Imprévisible	☐	☐	☐	☐	☐	☐	☐	Prévisible

---

*Les items précédés d'un astérisque sont des items inversés.

PeRDITA Questionnaire EN

Participant N° : Scenario N° :

In order to study your personal evaluation of each robot behavior, you're going to answer a questionnaire. You'll be asked to place a cross between two adjectives in the box that most closely matches your impression.

- Read carefully the bold statements before answering.
- You will be able to modify your answers later.
- There are no right or wrong answers. Answer as truthfully as possible.

Recall the properties of the scenario by ticking the appropriate boxes:

Execution Regime		Your objective	
Human-First ☐	Robot-First ☐	Finish the task as soon as possible ☐	Be freed as soon as possible ☐

In your opinion, the robot is rather:

Apathetic	☐	☐	☐	☐	☐	☐	☐	Responsive
*Competent	☐	☐	☐	☐	☐	☐	☐	Incompetent
Unintelligent	☐	☐	☐	☐	☐	☐	☐	Intelligent

In your opinion, generally, the interaction with the robot was:

Negative	☐	☐	☐	☐	☐	☐	☐	Positive
*Simple	☐	☐	☐	☐	☐	☐	☐	Complicated
*Clear	☐	☐	☐	☐	☐	☐	☐	Ambiguous

In your opinion, the collaboration with the robot to perform the task was:

Restrictive	☐	☐	☐	☐	☐	☐	☐	Adaptive
*Useful	☐	☐	☐	☐	☐	☐	☐	Useless
Inefficient	☐	☐	☐	☐	☐	☐	☐	Efficient

In your opinion, the robot choices of action were:

*Appropriate	☐	☐	☐	☐	☐	☐	☐	Inappropriate
Annoying	☐	☐	☐	☐	☐	☐	☐	Accommodating
Unpredictable	☐	☐	☐	☐	☐	☐	☐	Predictable

---

*Items preceded by an asterisk are reversed items.

N° Participant	Date	Age	Genre	Qu'elle est votre vision de la robotique ?	Affectation	Etes vous familier avec la robotique et la plannification de tâche ?	Avez vous déjà interagi avec un robot ? Si oui quel genre ? (Drone, Robot aspirateur, Humanoïde, Jouet)
1	17/12/2023 14:55:45	21	Male	4	EXT	Non	Tous ceux cités
2	17/12/2023 15:48:14	23	Woman	5	EXT	Non	pas vraiment
3	18/12/2023 10:02:21	26	Male	4	LAAS	Oui	PR2, Pepper
4	18/12/2023 10:45:39	25	Male	4	LAAS	Oui	thermomix
5	19/12/2023 10:40:40	28	Male	3	LAAS	Oui	Humanoïde
6	19/12/2023 11:42:17	26	Male	4	LAAS	Non	Drone
7	19/12/2023 14:42:18	26	Male	3	LAAS	Non	Pepper
8	19/12/2023 15:31:04	24	Woman	4	LAAS	Non	robot aspirateur
9	19/12/2023 19:06:39	26	Male	5	EXT	Non	Robot aspirateur, jouets
10	08/01/2024 10:28:39	30	Woman	4	LAAS	Oui	no
11	09/01/2024 16:02:02	24	Male	4	LAAS	Non	Pepper, PR2, jouets
12	09/01/2024 16:32:36	27	Male	5	LAAS	Non	Oui, tous
13	10/01/2024 16:25:59	27	Woman	4	LAAS	Non	électroménager + chatbox

Table A.1: Information on the participants. (part 1/2)

N° Participant	Date	Age	Genre	Qu'elle est votre vision de la robotique ?	Affectation	Etes vous familier avec la robotique et la plannification de tâche ?	Avez vous déjà interagi avec un robot ? Si oui quel genre ? (Drone, Robot aspirateur, Humanoïde, Jouet)
14	11/01/2024 16:00:19	25	Male	5	LAAS	Non	brat manipulateur
15	12/01/2024 10:11:42	23	Woman	5	LAAS	Non	oui (drone, jouet)
16	12/01/2024 10:56:29	30	Male	4	LAAS	Non	Humanoid
17	17/01/2024 09:27:52	61	Male	4	LAAS	Non	Oui
18	17/01/2024 10:37:37	26	Male	5	LAAS	Oui	Drone, Pepper, PR2, Chien robot, Aspirateur,tondeuse, ...
19	17/01/2024 17:23:21	27	Woman	4	EXT	Non	Non pas vraiment
20	17/01/2024 17:58:32	29	Male	5	EXT	Non	Drone, robot aspirateur
21	23/01/2024 10:26:45	25	Male	4	LAAS	Oui	Humanoïde, Robot mobile
22	23/01/2024 18:03:08	27	Woman	4	EXT	Non	Jouet
23	25/01/2024 12:15:45	47	Woman	4	LAAS	Oui	PR2, robot mobile, bras manipulateurs
24	25/01/2024 13:47:39	26	Woman	4	LAAS	Non	PR2 + Jouets
25	26/01/2024 14:20:54	62	Male	5	LAAS	Oui	Bras interactif, humanoïdes, drones

Table A.2: Information on the participants. (part 2/2)

Participant N°	S1 pose	S2 pose	S3 pose	S4 pose	S5 pose	S6 pose
1	1	2	3	6	5	4
2	3	1	2	6	5	4
3	5	3	6	1	2	4
4	2	5	4	6	3	1
5	1	5	2	6	4	3
6	6	3	4	2	1	5
7	3	5	6	4	1	2
8	4	5	3	6	2	1
9	4	2	1	5	6	3
10	1	5	3	2	4	6
11	5	4	6	2	3	1
12	3	1	5	2	6	4
13	1	2	5	6	4	3
14	4	2	3	1	5	6
15	4	1	6	5	3	2
16	2	6	3	1	4	5
17	2	3	1	5	4	6
18	5	1	3	6	2	4
19	6	1	5	2	4	3
20	4	5	6	3	1	2
21	6	2	4	3	1	5
22	1	4	6	3	2	5
23	2	4	1	3	6	5
24	5	1	4	3	2	6
25	2	1	4	6	3	5

Table A.3: Ordering in which each participant encountered each scenario.

N°	S1 Q1	S1 Q2	S1 Q3	S1 Q4	S1 Q5	S1 Q6	S1 Q7	S1 Q8	S1 Q9	S1 Q10	S1 Q11	S1 Q12	S2 Q1	S2 Q2	S2 Q3	S2 Q4	S2 Q5	S2 Q6	S2 Q7	S2 Q8	S2 Q9	S2 Q10	S2 Q11	S2 Q12
1	6	2	5	6	2	2	6	2	6	3	6	6	6	3	3	6	2	2	5	2	5	5	2	2
2	6	1	7	7	1	1	7	1	7	1	7	7	7	1	7	7	1	1	7	1	7	1	7	7
3	6	2	5	6	2	2	7	2	7	1	7	7	4	1	6	7	1	1	7	1	7	1	7	3
4	7	1	1	7	2	1	7	2	7	1	7	7	6	1	1	7	1	3	7	1	7	1	7	5
5	6	3	3	6	3	3	6	3	6	2	5	6	6	2	5	6	3	3	6	3	6	3	5	6
6	4	1	7	7	1	1	7	1	7	1	7	7	7	1	7	7	1	2	7	1	7	1	7	6
7	6	2	2	6	2	2	5	3	5	2	5	5	6	2	2	6	2	2	6	2	6	2	4	4
8	7	1	7	7	1	1	7	1	7	1	7	7	7	2	7	7	3	2	6	3	6	2	6	5
9	7	1	7	7	1	1	7	1	7	1	7	7	7	1	7	7	2	1	6	1	7	1	7	7
10	6	2	5	6	2	3	6	2	5	2	6	6	6	2	6	6	2	2	6	2	6	2	6	6
11	7	1	7	7	1	1	7	1	7	1	7	7	7	1	7	7	1	1	7	1	7	1	7	7
12	6	2	6	5	2	2	6	3	6	2	6	6	3	3	3	4	5	6	3	3	2	5	3	4
13	5,5	6	4	6	1	1	6	3	5	2	7	7	7	1	4	7	1	1	7	2	7	1	7	6
14	6	2	3	5	3	2	4	3	6	2	6	4	6	2	6	6	3	3	6	2	6	3	6	5
15	7	1	7	7	1	1	7	1	7	1	7	7	7	1	7	7	1	1	6	1	7	1	7	7
16	6	3	6	6	2	2	7	6	6	2	6	6	6	2	6	6	2	2	7	2	7	2	5	5
17	6	2	5	6	3	2	6	6	6	2	5	5	5	2	5	6	2	3	4	1	6	2	4	6
18	7	2	6	6	2	2	7	1	6	1	7	6	6	2	5	6	2	2	6	2	6	1	6	7
19	6	1	7	6	1	1	7	1	7	1	7	7	7	1	6	6	4	4	6	2	6	1	7	7
20	7	1	7	7	1	1	7	1	7	1	7	7	7	1	7	7	1	1	7	1	7	1	7	6
21	6	2	6	6	2	2	6	2	6	2	6	6	7	2	6	6	2	1	6	2	6	2	7	6
22	7	1	4	7	1	1	7	1	7	1	7	7	7	2	4	6	2	2	5	1	7	1	7	6
23	6	2	5	6	2	2	6	2	6	2	6	6	6	2	5	6	2	2	5	3	5	2	6	6
24	7	1	7	7	1	1	7	1	7	1	7	7	7	1	7	7	1	1	7	1	7	7	7	4
25	6	2	6	5	2	2	6	2	6	2	6	6	6	2	6	5	2	2	3	2	5	2	6	2

Table A.4: Participants' answers to the twelve questions of the questionnaire (from Q1 to Q12), for each of the six scenarios (from S1 to S6), on a - Part 1

N°	S3 Q1	S3 Q2	S3 Q3	S3 Q4	S3 Q5	S3 Q6	S3 Q7	S3 Q8	S3 Q9	S3 Q10	S3 Q11	S3 Q12	S4 Q1	S4 Q2	S4 Q3	S4 Q4	S4 Q5	S4 Q6	S4 Q7	S4 Q8	S4 Q9	S4 Q10	S4 Q11	S4 Q12
1	6	3	3	6	2	2	5	2	5	5	2	2	6	2	6	6	2	2	6	1	6	2	6	5
2	6	1	7	7	1	1	7	1	7	1	7	7	7	6	3	6	6	7	2	3	3	6	3	6
3	6	2	6	5	5	2	5	2	6	1	6	7	4	7	2	2	1	2	1	6	1	7	2	1
4	7	1	1	6	1	1	7	1	7	2	7	7	7	7	1	3	1	1	3	2	4	6	3	7
5	6	2	6	5	2	2	6	4	7	2	7	7	4	6	1	1	6	7	1	7	1	7	1	1
6	7	1	6	7	2	2	7	1	6	1	7	7	7	6	2	1	7	2	1	4	2	6	1	2
7	4	2	6	6	2	2	4	5	6	2	6	6	6	2	6	2	6	6	2	6	2	6	2	6
8	7	1	7	7	1	1	7	1	7	1	7	7	5	3	5	5	2	5	3	5	3	2	5	3
9	7	1	7	7	2	1	6	1	7	1	7	7	7	6	2	5	1	1	3	4	2	6	2	1
10	5	3	5	6	2	3	5	3	5	5	4	4	3	5	3	3	2	4	3	4	2	6	2	4
11	7	1	7	7	1	1	7	1	7	1	7	7	7	1	6	6	1	1	5	1	5	2	7	6
12	6	3	5	5	1	2	4	3	5	1	6	7	2	3	4	2	3	5	3	6	3	5	1	1
13	7	1	6	7	1	1	7	2	7	1	7	7	7	2	3	6	1	1	3	4	5	5	6	3
14	5	3	3	6	2	3	5	3	6	3	6	5	5	5	5	5	2	3	5	3	4	4	4	6
15	7	2	6	5	3	4	7	1	7	2	7	7	7	4	6	3	4	4	2	1	5	2	3	6
16	6	3	6	6	2	2	7	6	6	2	6	6	5	5	5	5	3	3	4	4	3	5	3	3
17	5	6	3	3	5	4	3	3	1	6	3	4	5	6	2	3	5	5	3	5	6	6	3	2
18	7	2	6	7	1	2	6	1	6	2	6	6	2	6	2	3	6	7	1	7	2	7	1	1
19	7	1	7	7	1	1	7	1	7	1	7	7	6	7	1	1	7	5	1	7	2	7	1	1
20	7	1	7	7	1	1	7	1	7	1	7	7	4	5	3	3	5	5	3	3	4	3	5	4
21	6	2	6	6	2	2	6	2	6	2	6	6	5	4	3	3	5	5	2	5	2	5	3	3
22	7	1	4	7	1	1	7	1	7	1	7	7	7	6	3	1	6	7	2	6	3	7	1	1
23	6	2	6	6	2	2	6	2	6	2	6	6	6	3	5	5	3	3	6	2	6	2	5	6
24	7	1	6	7	1	1	7	1	7	1	7	6	7	3	2	6	5	1	5	4	4	5	3	4
25	6	2	2	7	2	2	6	2	6	2	7	6	6	6	2	2	4	2	3	3	2	6	2	3

Table A.5: Participants' answers to the 12 questions (Q) of the questionnaire, for each of the 6 scenarios (S) - Part 2

N°	S5 Q1	S5 Q2	S5 Q3	S5 Q4	S5 Q5	S5 Q6	S5 Q7	S5 Q8	S5 Q9	S5 Q10	S5 Q11	S5 Q12	S6 Q1	S6 Q2	S6 Q3	S6 Q4	S6 Q5	S6 Q6	S6 Q7	S6 Q8	S6 Q9	S6 Q10	S6 Q11	S6 Q12
1	5	6	2	4	2	2	3	5	3	7	3	3	6	3	2	5	2	2	3	6	3	6	1	1
2	7	1	7	7	1	2	7	1	7	1	7	7	7	5	4	5	1	5	4	2	5	5	3	5
3	6	2	4	6	2	1	5	3	7	1	7	4	5	3	3	3	4	4	2	4	1	5	3	5
4	7	1	1	6	1	2	7	1	6	1	7	7	5	1	5	1	2	2	3	4	6	4	6	7
5	4	3	5	4	1	2	4	4	4	3	4	6	6	3	3	2	4	5	1	6	2	6	1	2
6	7	1	7	7	1	2	6	1	6	3	5	5	7	6	5	2	6	4	1	3	4	6	1	4
7	6	2	6	5	4	5	4	3	6	5	4	4	6	2	6	2	6	6	3	3	5	5	2	2
8	7	1	6	7	1	1	7	1	7	2	7	6	6	3	5	5	5	6	3	3	2	6	3	6
9	7	1	6	7	1	2	7	1	6	2	7	6	5	2	3	6	2	5	7	1	3	5	3	3
10	6	1	7	6	2	2	6	1	6	3	5	5	5	4	2	3	2	4	2	5	2	5	2	4
11	7	1	7	7	1	1	7	1	7	2	7	6	7	1	6	3	1	1	3	2	3	5	3	5
12	5	3	2	3	1	5	7	2	5	4	3	5	5	6	3	1	2	6	1	6	2	7	2	3
13	5	2	4	6	1	1	6	1	6	2	7	5	5	4	4	5	1	1	5	4	5	5	4	5
14	6	2	2	6	2	2	6	2	6	2	6	5	5	5	5	4	2	2	5	5	3	4	4	4
15	7	2	5	6	3	4	7	2	6	3	7	4	7	3	3	3	2	4	2	5	2	5	2	3
16	7	2	7	6	2	2	5	2	5	2	6	6	4	4	4	5	3	4	4	3	3	5	3	5
17	5	4	5	4	4	5	4	2	3	5	4	3	5	3	5	2	6	4	4	3	2	6	2	3
18	6	3	6	7	3	3	6	1	6	3	6	7	6	2	5	6	2	2	5	2	2	6	2	6
19	6	4	5	5	3	3	4	2	5	6	4	4	5	2	5	4	4	5	5	2	6	4	2	6
20	7	1	7	7	1	1	7	1	6	1	7	6	6	2	6	4	3	2	6	2	6	3	6	6
21	6	2	6	6	2	2	6	2	6	2	6	6	6	2	6	4	3	2	5	3	4	3	3	4
22	7	1	4	7	1	1	6	1	7	2	7	6	7	2	6	5	2	2	5	3	4	3	4	1
23	6	3	5	6	3	3	4	2	5	3	7	5	5	3	5	5	3	4	5	2	5	3	6	5
24	7	2	5	7	1	1	7	1	6	2	7	7	5	3	5	2	2	5	1	7	1	7	1	4
25	6	2	2	7	2	2	7	1	7	2	6	6	6	2	2	4	4	3	2	3	2	6	2	2

Table A.6: Participants' answers to the 12 questions (Q) of the questionnaire, for each of the 6 scenarios (S) - Part 3

M1	task_completion_time	M17	h_action_time_average
M2	number_steps	M18	h_action_time_sd
M3	nb_h_optimal_action	M19	h_action_time_max
M4	ratio_h_optimal_action	M20	h_action_time_min
M5	decision_time_total	M21	r_action_nb
M6	decision_time_average	M22	r_action_time_total
M7	decision_time_sd	M23	r_action_time_average
M8	decision_time_max	M24	r_action_time_sd
M9	decision_time_min	M25	r_action_time_max
M10	wait_ns_total	M26	r_action_time_min
M11	wait_ns_average	M27	time_human_free
M12	wait_ns_sd	M28	plan_mvt_total
M13	wait_ns_max	M29	plan_mvt_min
M14	wait_ns_min	M30	plan_mvt_max
M15	h_action_nb	M31	plan_mvt_average
M16	h_action_time_total	M32	plan_mvt_sd

Table A.7: Execution metrics abbreviations.

N°	S1 M1	S1 M2	S1 M3	S1 M4	S1 M5	S1 M6	S1 M7	S1 M8	S1 M9	S1 M10	S1 M11	S1 M12	S1 M13	S1 M14	S1 M15	S1 M16	S1 M17	S1 M18	S1 M19
1	56,6	10,0	10,0	100,0	8,2	1,0	0,5	1,8	0,3	12,5	1,6	0,9	3,2	0,6	8,0	26,2	3,3	0,6	4,4
2	58,7	10,0	10,0	100,0	9,7	1,2	0,5	2,0	0,6	14,3	1,8	1,2	3,9	0,7	8,0	25,6	3,2	0,6	4,2
3	61,6	10,0	9,0	90,0	10,3	1,0	0,8	2,9	0,4	14,9	1,9	1,2	4,0	0,7	8,0	26,6	3,3	0,6	4,3
4	56,7	10,0	9,0	90,0	3,8	0,4	0,4	1,2	0,0	17,6	2,2	1,5	5,3	0,6	8,0	26,1	3,3	0,6	4,3
5	55,3	10,0	10,0	100,0	2,9	0,4	0,5	1,2	0,0	17,7	2,2	1,4	4,3	0,6	8,0	26,3	3,3	0,6	4,3
6	52,6	10,0	10,0	100,0	3,6	0,4	0,5	1,7	0,0	14,4	1,8	1,0	3,8	0,7	8,0	25,9	3,2	0,6	4,3
7	83,5	14,0	10,0	71,4	12,0	0,9	0,7	2,4	0,1	11,8	2,0	1,1	4,0	0,6	6,0	18,7	3,1	0,5	4,2
8	60,6	10,0	9,0	90,0	2,8	0,3	0,6	1,9	0,0	14,5	1,8	1,2	4,2	0,7	8,0	26,4	3,3	0,6	4,3
9	61,3	10,0	10,0	100,0	7,3	0,9	0,5	2,0	0,6	18,3	2,3	1,3	4,3	0,7	8,0	25,9	3,2	0,6	4,3
10	62,4	10,0	10,0	100,0	14,1	1,8	2,3	6,8	0,0	13,9	1,7	1,0	3,6	0,7	8,0	26,0	3,3	0,6	4,3
11	58,6	10,0	10,0	100,0	4,6	0,5	0,5	1,8	0,1	20,2	2,5	1,8	6,5	0,6	8,0	25,8	3,2	0,6	4,3
12	58,3	10,0	10,0	100,0	3,6	0,5	1,0	3,2	0,0	18,9	2,4	1,7	5,7	0,7	8,0	26,2	3,3	0,6	4,3
13	61,4	12,0	10,0	83,3	1,6	0,1	0,2	0,4	0,0	17,9	1,8	1,1	4,1	0,7	10,0	32,7	3,3	0,6	4,3
14	57,5	10,0	10,0	100,0	3,5	0,4	0,6	1,7	0,0	20,7	2,6	1,8	6,8	0,6	8,0	26,0	3,2	0,6	4,3
15	62,3	10,0	9,0	90,0	2,7	0,3	0,6	1,9	0,0	18,3	2,3	1,5	4,8	0,6	8,0	26,4	3,3	0,6	4,3
16	60,2	10,0	10,0	100,0	4,5	0,6	1,0	3,1	0,0	19,4	2,4	1,5	4,7	0,6	8,0	25,9	3,2	0,6	4,3
17	59,3	10,0	10,0	100,0	10,2	1,3	1,8	5,8	0,0	13,0	1,6	0,9	3,0	0,6	8,0	25,8	3,2	0,6	4,2
18	56,5	10,0	10,0	100,0	2,7	0,3	0,2	0,8	0,0	18,9	2,4	1,7	5,6	0,6	8,0	26,1	3,3	0,6	4,3
19	59,7	10,0	9,0	90,0	4,9	0,5	1,0	3,0	0,0	14,2	1,8	1,5	5,1	0,6	8,0	26,6	3,3	0,6	4,3
20	54,8	10,0	10,0	100,0	2,8	0,3	0,2	0,7	0,0	17,2	2,2	1,7	5,1	0,6	8,0	26,2	3,3	0,6	4,4
21	55,2	10,0	10,0	100,0	1,6	0,2	0,2	0,5	0,0	18,6	2,3	1,4	4,9	0,6	8,0	25,8	3,2	0,6	4,3
22	60,3	10,0	10,0	100,0	8,7	1,1	0,5	1,9	0,5	16,1	2,0	1,0	4,0	0,6	8,0	26,1	3,3	0,6	4,3
23	67,4	12,0	10,0	83,3	6,8	0,7	1,1	3,5	0,0	19,5	2,0	1,4	4,6	0,7	10,0	32,6	3,3	0,6	4,3
24	56,5	10,0	10,0	100,0	9,8	1,1	0,5	2,0	0,4	11,6	1,5	0,7	2,5	0,7	8,0	26,0	3,3	0,6	4,3
25	58,5	10,0	10,0	100,0	2,7	0,3	0,5	1,8	0,0	18,9	2,4	1,4	5,0	0,6	8,0	25,8	3,2	0,6	4,2

Table A.8: Execution metrics - Part 1

N°	S1 M20	S1 M21	S1 M22	S1 M23	S1 M24	S1 M25	S1 M26	S1 M27	S1 M28	S1 M29	S1 M30	S1 M31	S1 M32	S2 M1	S2 M2	S2 M3	S2 M4	S2 M5	S2 M6
1	2,6	8,0	28,2	3,5	0,6	4,5	2,6	22,1	5,7	0,0	1,2	0,6	0,3	59,5	10,0	10,0	100,0	18,3	2,3
2	2,5	8,0	31,3	3,9	0,9	5,8	2,7	22,7	3,2	0,0	0,7	0,3	0,2	54,7	10,0	10,0	100,0	7,8	1,0
3	2,6	8,0	29,8	3,7	0,9	6,1	3,0	61,6	5,8	0,1	1,8	0,7	0,4	53,9	10,0	10,0	100,0	11,0	1,4
4	2,5	8,0	33,4	4,2	1,2	7,0	2,9	56,7	4,1	0,2	0,7	0,5	0,2	50,0	10,0	10,0	100,0	3,8	0,4
5	2,7	8,0	34,0	4,2	1,4	6,7	3,1	18,8	3,9	0,0	0,7	0,4	0,3	51,1	10,0	10,0	100,0	1,6	0,2
6	2,7	8,0	28,8	3,6	0,8	5,2	2,8	18,2	5,5	0,0	0,8	0,6	0,3	50,0	10,0	10,0	100,0	2,0	0,3
7	2,6	12,0	53,2	4,4	1,5	7,9	3,1	21,0	6,1	0,0	0,7	0,4	0,3	54,4	10,0	10,0	100,0	6,5	0,8
8	2,6	8,0	32,5	4,1	1,2	5,8	2,7	60,6	5,0	0,3	1,0	0,6	0,2	59,4	11,0	10,0	90,9	6,2	0,8
9	2,6	8,0	35,4	4,4	1,0	5,9	3,0	22,9	4,2	0,0	0,8	0,4	0,3	52,7	10,0	10,0	100,0	3,8	0,5
10	2,6	8,0	30,6	3,8	0,8	5,0	2,6	25,2	4,5	0,0	0,7	0,5	0,3	64,0	12,0	10,0	83,3	17,1	2,1
11	2,6	8,0	34,6	4,3	1,3	6,9	3,1	22,5	5,4	0,0	1,1	0,5	0,3	53,1	10,0	10,0	100,0	9,7	1,2
12	2,6	8,0	35,3	4,4	1,2	6,2	3,0	21,0	4,9	0,0	1,2	0,5	0,4	54,5	10,0	10,0	100,0	12,3	1,5
13	2,6	8,0	32,2	4,0	0,8	5,6	3,3	17,5	4,9	0,0	0,8	0,4	0,3	57,0	10,0	10,0	100,0	8,2	1,0
14	2,7	8,0	34,4	4,3	1,2	6,2	3,0	24,4	5,0	0,0	2,3	0,5	0,6	63,0	11,0	10,0	90,9	12,7	1,6
15	2,7	8,0	33,9	4,2	1,2	6,5	3,0	62,3	5,2	0,5	0,8	0,6	0,1	58,3	10,0	10,0	100,0	12,0	1,5
16	2,6	8,0	36,2	4,5	1,2	6,2	2,8	19,8	5,0	0,0	0,7	0,5	0,3	60,9	11,0	10,0	90,9	11,3	1,4
17	2,6	8,0	30,6	3,8	1,1	5,5	2,3	19,2	3,7	0,0	0,7	0,4	0,3	75,8	14,0	10,0	71,4	5,4	0,9
18	2,6	8,0	34,3	4,3	1,5	7,4	3,1	18,4	5,0	0,0	0,8	0,5	0,3	52,3	10,0	10,0	100,0	5,8	0,6
19	2,6	8,0	29,9	3,7	1,2	6,7	2,9	59,7	4,3	0,1	0,7	0,5	0,2	54,0	11,0	10,0	90,9	0,9	0,1
20	2,7	8,0	33,0	4,1	1,3	6,7	3,1	20,3	4,4	0,0	0,8	0,4	0,3	56,7	10,0	10,0	100,0	10,9	1,4
21	2,6	8,0	33,9	4,2	1,0	6,0	2,7	20,5	5,0	0,0	0,9	0,5	0,3	44,7	10,0	10,0	100,0	1,7	0,2
22	2,7	8,0	31,3	3,9	0,9	5,8	2,8	23,2	5,7	0,0	0,9	0,6	0,3	51,8	10,0	10,0	100,0	9,0	1,1
23	2,6	8,0	32,2	4,0	0,9	5,4	3,1	24,5	5,3	0,0	0,7	0,4	0,3	64,6	12,0	10,0	83,3	6,9	1,0
24	2,6	8,0	28,3	3,5	0,4	4,1	2,8	20,9	4,1	0,0	0,7	0,4	0,3	54,2	10,0	10,0	100,0	10,2	1,3
25	2,6	8,0	36,5	4,6	1,3	6,6	2,7	19,9	4,6	0,0	0,9	0,5	0,3	65,2	12,0	10,0	83,3	11,5	1,4

Table A.9: Execution metrics - Part 2

N°	S2 M7	S2 M8	S2 M9	S2 M10	S2 M11	S2 M12	S2 M13	S2 M14	S2 M15	S2 M16	S2 M17	S2 M18	S2 M19	S2 M20	S2 M21	S2 M22	S2 M23	S2 M24	S2 M25
1	1,4	5,4	0,9	5,9	0,7	0,1	1,1	0,6	8,0	26,0	3,3	0,6	4,3	2,6	8,0	31,2	3,9	0,9	5,9
2	0,4	1,9	0,4	8,3	1,0	0,4	1,6	0,6	8,0	26,1	3,3	0,6	4,3	2,6	8,0	31,2	3,9	0,6	4,9
3	0,8	2,8	0,2	8,3	1,0	0,5	2,0	0,6	8,0	26,2	3,3	0,6	4,3	2,6	8,0	30,3	3,8	0,8	5,3
4	0,5	1,5	0,0	11,7	1,5	1,1	3,6	0,6	8,0	26,2	3,3	0,6	4,2	2,6	8,0	30,1	3,8	1,2	6,7
5	0,2	0,6	0,0	14,8	1,8	1,5	5,3	0,6	8,0	26,3	3,3	0,6	4,3	2,7	8,0	32,1	4,0	1,2	6,9
6	0,2	0,6	0,0	12,9	1,6	1,2	4,3	0,7	8,0	26,1	3,3	0,6	4,3	2,6	8,0	32,6	4,1	0,9	5,7
7	1,5	4,7	0,0	10,8	1,4	1,0	3,8	0,6	8,0	26,3	3,3	0,6	4,3	2,6	8,0	31,9	4,0	0,9	5,8
8	1,2	3,8	0,0	12,4	1,6	1,5	4,8	0,6	8,0	26,3	3,3	0,6	4,3	2,7	8,0	35,4	4,4	1,2	6,4
9	0,5	1,6	0,0	11,1	1,4	0,8	2,5	0,6	8,0	26,2	3,3	0,6	4,3	2,6	8,0	33,0	4,1	1,2	6,3
10	1,9	5,9	0,0	8,2	1,0	0,6	2,4	0,6	8,0	26,1	3,3	0,6	4,3	2,6	8,0	30,2	3,8	1,1	5,9
11	0,9	2,9	0,0	9,7	1,2	0,8	2,6	0,6	8,0	25,9	3,2	0,6	4,3	2,6	8,0	30,0	3,8	1,1	6,3
12	1,1	3,9	0,0	6,2	0,8	0,2	1,2	0,6	8,0	26,1	3,3	0,7	4,4	2,6	8,0	30,9	3,9	0,8	5,4
13	1,2	3,7	0,0	14,4	1,8	1,3	4,7	0,6	8,0	26,0	3,3	0,6	4,3	2,6	8,0	34,8	4,3	2,0	8,7
14	1,9	4,8	0,0	9,7	1,2	0,6	2,4	0,6	8,0	26,0	3,3	0,6	4,4	2,6	8,0	31,4	3,9	0,6	4,8
15	1,5	4,7	0,0	12,2	1,5	1,6	5,6	0,6	8,0	26,0	3,2	0,6	4,3	2,7	8,0	34,5	4,3	1,2	6,3
16	1,3	3,5	0,0	7,6	0,9	0,7	2,8	0,6	8,0	26,3	3,3	0,7	4,4	2,7	8,0	34,3	4,3	1,1	6,2
17	1,8	5,0	0,0	13,7	2,3	1,0	3,6	0,7	6,0	19,7	3,3	0,7	4,3	2,6	12,0	48,6	4,1	1,0	6,0
18	0,8	2,5	0,0	11,5	1,4	0,9	3,3	0,6	8,0	26,5	3,3	0,6	4,4	2,6	8,0	31,1	3,9	0,7	4,7
19	0,2	0,6	0,0	8,9	1,1	0,6	2,2	0,6	8,0	26,3	3,3	0,6	4,3	2,6	8,0	28,1	3,5	0,3	3,9
20	1,4	4,7	0,0	10,6	1,3	1,4	5,0	0,6	8,0	26,4	3,3	0,6	4,3	2,7	8,0	33,1	4,1	0,9	5,8
21	0,3	0,9	0,0	9,0	1,1	0,5	1,9	0,6	8,0	26,1	3,3	0,6	4,3	2,6	8,0	26,4	3,3	0,3	3,7
22	1,4	4,2	0,0	8,0	1,0	0,6	2,4	0,6	8,0	26,3	3,3	0,6	4,3	2,7	8,0	32,1	4,0	0,8	5,5
23	1,0	2,9	0,0	6,6	1,1	0,6	2,1	0,6	6,0	20,4	3,4	0,7	4,4	2,6	10,0	40,5	4,1	0,6	4,8
24	0,9	2,9	0,4	8,1	1,0	0,8	3,2	0,6	8,0	26,3	3,3	0,6	4,3	2,6	8,0	31,4	3,9	0,9	5,4
25	1,6	4,7	0,0	8,5	1,1	0,8	3,1	0,6	8,0	26,0	3,3	0,6	4,4	2,6	8,0	36,5	4,6	1,3	7,2

Table A.10: Execution metrics - Part 3

N°	S2 M26	S2 M27	S2 M28	S2 M29	S2 M30	S2 M31	S2 M32	S3 M1	S3 M2	S3 M3	S3 M4	S3 M5	S3 M6	S3 M7	S3 M8	S3 M9	S3 M10	S3 M11	S3 M12
1	2,8	26,0	5,1	0,2	1,1	0,6	0,3	54,8	10,0	10,0	100,0	4,9	0,6	0,2	1,0	0,4	15,4	1,9	1,4
2	2,9	54,7	5,2	0,5	0,7	0,6	0,1	55,0	10,0	10,0	100,0	6,3	0,8	0,5	2,0	0,0	14,6	1,8	0,9
3	2,8	22,9	4,7	0,2	0,7	0,6	0,2	88,1	12,0	10,0	83,3	9,0	1,1	0,6	2,1	0,1	16,6	2,1	1,6
4	2,7	50,0	4,1	0,3	0,7	0,5	0,2	60,7	10,0	9,0	90,0	4,6	0,5	0,7	2,5	0,0	20,3	2,5	1,7
5	3,0	18,8	5,1	0,3	1,2	0,6	0,3	60,4	12,0	10,0	83,3	0,7	0,1	0,1	0,5	0,0	19,2	1,9	1,1
6	3,0	19,9	4,7	0,2	1,0	0,6	0,2	83,3	12,0	10,0	83,3	6,4	0,7	0,6	2,2	0,0	19,5	2,4	1,6
7	3,0	54,4	5,8	0,2	1,5	0,7	0,3	84,3	12,0	10,0	83,3	20,5	1,9	3,0	10,8	0,1	22,4	2,2	1,6
8	2,9	23,0	4,9	0,3	0,9	0,6	0,2	63,9	12,0	10,0	83,3	1,9	0,2	0,4	1,2	0,0	21,1	2,1	1,2
9	2,6	19,1	5,2	0,4	0,8	0,7	0,1	63,5	10,0	10,0	100,0	6,8	0,9	0,5	2,0	0,4	22,2	2,8	1,8
10	2,7	24,2	4,5	0,4	0,7	0,6	0,1	80,1	14,0	11,0	78,6	2,4	0,3	0,3	1,0	0,0	17,5	2,2	2,0
11	2,8	20,6	5,3	0,3	1,0	0,7	0,2	58,6	10,0	10,0	100,0	3,0	0,3	0,2	0,5	0,0	21,5	2,7	1,9
12	3,1	19,6	4,8	0,2	0,7	0,6	0,2	62,8	12,0	10,0	83,3	2,9	0,3	0,4	1,2	0,0	17,1	1,7	1,0
13	2,9	21,4	5,8	0,3	1,4	0,7	0,3	71,9	12,0	11,0	91,7	1,2	0,1	0,2	0,5	0,0	20,2	2,5	1,7
14	2,8	25,6	5,1	0,4	0,9	0,6	0,1	71,1	10,0	10,0	100,0	15,2	1,9	1,4	3,8	0,1	21,6	2,7	1,8
15	2,9	25,5	5,4	0,1	1,3	0,7	0,4	74,1	12,0	11,0	91,7	3,8	0,6	0,8	1,7	0,0	14,8	2,5	1,9
16	2,7	21,6	6,4	0,2	1,5	0,8	0,4	67,8	12,0	10,0	83,3	5,1	0,6	1,1	3,1	0,0	12,8	1,6	1,4
17	2,7	32,0	8,3	0,3	1,3	0,7	0,2	116,2	17,0	14,0	82,4	0,8	0,1	0,1	0,4	0,0	11,0	1,8	1,2
18	3,0	52,3	4,7	0,2	0,7	0,6	0,2	53,5	10,0	10,0	100,0	1,7	0,2	0,3	0,9	0,0	18,1	2,3	1,1
19	3,0	54,0	4,3	0,2	0,7	0,5	0,2	55,5	10,0	10,0	100,0	1,3	0,2	0,2	0,5	0,0	17,9	2,2	1,5
20	2,9	23,5	5,2	0,2	1,2	0,7	0,3	55,5	10,0	10,0	100,0	2,6	0,3	0,4	1,3	0,0	17,7	2,2	1,7
21	2,7	16,6	4,7	0,3	0,7	0,6	0,1	51,7	10,0	10,0	100,0	1,6	0,2	0,2	0,6	0,0	15,2	1,9	1,0
22	2,9	18,1	3,8	0,2	0,8	0,5	0,2	51,8	10,0	10,0	100,0	2,2	0,3	0,2	0,8	0,0	16,1	2,0	1,4
23	2,8	64,6	6,2	0,1	1,4	0,6	0,3	61,1	12,0	10,0	83,3	0,4	0,0	0,0	0,1	0,0	19,5	2,0	1,2
24	2,8	23,0	4,1	0,2	0,7	0,5	0,2	65,7	10,0	10,0	100,0	8,7	1,1	1,1	3,7	0,2	19,7	2,5	1,4
25	3,0	20,8	4,3	0,3	0,8	0,5	0,2	53,7	10,0	10,0	100,0	1,7	0,2	0,4	1,3	0,0	16,8	2,1	1,1

Table A.11: Execution metrics - Part 4

N°	S3 M13	S3 M14	S3 M15	S3 M16	S3 M17	S3 M18	S3 M19	S3 M20	S3 M21	S3 M22	S3 M23	S3 M24	S3 M25	S3 M26	S3 M27	S3 M28	S3 M29	S3 M30	S3 M31
1	4,8	0,6	8,0	25,9	3,2	0,6	4,3	2,6	8,0	30,1	3,8	0,9	5,8	3,1	21,6	4,3	0,0	0,7	0,4
2	3,1	0,6	8,0	25,8	3,2	0,6	4,3	2,6	8,0	27,6	3,4	0,6	4,4	2,6	21,2	5,8	0,0	1,1	0,6
3	5,6	0,7	8,0	26,0	3,3	0,5	4,2	2,6	10,0	40,5	4,1	1,2	6,5	2,2	88,1	5,0	0,2	0,7	0,5
4	6,2	0,7	8,0	26,4	3,3	0,6	4,3	2,6	8,0	33,4	4,2	1,4	7,1	3,0	60,7	6,2	0,3	1,9	0,8
5	3,6	0,6	10,0	32,8	3,3	0,6	4,3	2,6	8,0	31,7	4,0	0,7	4,9	2,8	19,2	4,0	0,0	0,7	0,3
6	6,2	0,7	8,0	26,1	3,3	0,5	4,2	2,6	10,0	41,8	4,2	1,2	7,6	3,2	83,3	5,5	0,2	0,7	0,6
7	5,0	0,6	10,0	32,5	3,3	0,6	4,2	2,7	10,0	45,7	4,6	1,3	7,5	2,9	32,9	5,8	0,0	0,8	0,5
8	3,5	0,7	10,0	32,9	3,3	0,6	4,3	2,6	8,0	33,2	4,2	0,8	5,5	3,0	21,2	4,5	0,0	0,7	0,4
9	5,8	0,7	8,0	26,1	3,3	0,6	4,3	2,6	8,0	36,3	4,5	1,5	7,4	3,1	25,4	4,7	0,0	1,1	0,5
10	6,5	0,6	8,0	25,8	3,2	0,5	4,2	2,7	10,0	44,2	4,4	1,5	8,3	2,8	32,8	5,3	0,0	1,0	0,4
11	6,1	0,6	8,0	25,9	3,2	0,6	4,3	2,6	8,0	36,4	4,5	1,5	7,7	3,0	21,6	4,0	0,0	1,0	0,4
12	3,4	0,6	10,0	32,9	3,3	0,6	4,3	2,6	8,0	31,4	3,9	0,4	4,6	3,3	19,2	4,4	0,0	0,7	0,4
13	6,1	0,7	8,0	25,7	3,2	0,5	4,2	2,7	10,0	42,8	4,3	1,5	7,4	2,6	20,2	6,1	0,0	0,7	0,5
14	6,3	0,7	8,0	26,0	3,2	0,6	4,3	2,6	8,0	35,0	4,4	1,2	7,0	3,1	25,8	4,9	0,0	1,0	0,5
15	6,2	0,7	6,0	19,2	3,2	0,5	4,3	2,7	10,0	45,2	4,5	1,4	7,5	2,9	36,0	5,3	0,0	0,7	0,4
16	5,0	0,6	8,0	26,6	3,3	0,6	4,3	2,6	8,0	31,1	3,9	1,0	6,1	2,9	21,4	4,1	0,0	0,7	0,3
17	3,7	0,6	6,0	19,8	3,3	0,7	4,3	2,6	14,0	61,0	4,4	1,2	7,1	2,7	31,5	9,2	0,0	1,2	0,5
18	3,9	0,6	8,0	26,1	3,3	0,6	4,2	2,6	8,0	31,5	3,9	0,8	5,2	2,7	21,1	4,8	0,0	0,8	0,5
19	5,2	0,6	8,0	26,2	3,3	0,6	4,3	2,6	8,0	34,0	4,2	1,4	6,6	2,8	19,6	4,3	0,0	0,7	0,4
20	6,1	0,7	8,0	26,5	3,3	0,6	4,4	2,7	8,0	32,3	4,0	1,3	7,2	2,9	19,6	5,0	0,0	0,7	0,5
21	3,7	0,6	8,0	26,1	3,3	0,6	4,3	2,6	8,0	29,1	3,6	0,5	4,7	3,2	19,0	5,0	0,0	0,7	0,5
22	5,2	0,6	8,0	26,2	3,3	0,6	4,3	2,6	8,0	28,9	3,6	1,2	6,6	2,6	18,1	4,9	0,0	0,8	0,5
23	4,1	0,6	10,0	33,0	3,3	0,6	4,3	2,6	8,0	31,8	4,0	0,9	6,0	3,0	18,6	4,7	0,0	0,7	0,4
24	4,5	0,7	8,0	26,2	3,3	0,6	4,4	2,6	8,0	36,0	4,5	1,0	5,5	2,9	27,7	4,9	0,0	1,2	0,5
25	4,3	0,7	8,0	25,8	3,2	0,6	4,3	2,6	8,0	31,2	3,9	0,5	5,3	3,5	18,4	4,8	0,0	0,7	0,5

Table A.12: Execution metrics - Part 5

N°	S3 M32	S4 M1	S4 M2	S4 M3	S4 M4	S4 M5	S4 M6	S4 M7	S4 M8	S4 M9	S4 M10	S4 M11	S4 M12	S4 M13	S4 M14	S4 M15	S4 M16	S4 M17	S4 M18
1	0,2	91,7	18,0	16,0	88,9	0,6	0,3	0,0	0,3	0,3	1,3	0,7	0,0	0,7	0,6	2,0	6,0	3,0	0,3
2	0,3	80,7	16,0	16,0	100,0	0,7	0,2	0,1	0,3	0,0	6,0	1,5	1,4	4,0	0,7	4,0	12,9	3,2	0,7
3	0,2	82,4	16,0	16,0	100,0	3,7	0,9	0,7	2,1	0,1	7,2	1,8	2,0	5,2	0,6	4,0	13,1	3,3	0,6
4	0,4	78,1	16,0	16,0	100,0	1,8	0,4	0,4	0,9	0,0	4,4	1,1	0,7	2,4	0,7	4,0	13,4	3,3	0,7
5	0,3	78,2	16,0	16,0	100,0	2,9	0,7	0,7	1,8	0,0	5,2	1,3	1,1	3,1	0,6	4,0	13,0	3,3	0,6
6	0,2	79,0	16,0	16,0	100,0	0,4	0,1	0,1	0,3	0,0	5,6	1,4	1,3	3,6	0,6	4,0	13,0	3,3	0,7
7	0,3	86,3	16,0	16,0	100,0	4,0	1,0	1,2	2,9	0,0	3,4	0,8	0,3	1,4	0,6	4,0	13,0	3,3	0,7
8	0,3	77,8	16,0	16,0	100,0	2,9	0,7	1,1	2,6	0,0	8,8	2,2	2,2	6,0	0,6	4,0	13,1	3,3	0,6
9	0,3	88,0	16,0	16,0	100,0	1,6	0,4	0,7	1,6	0,0	8,0	2,0	2,2	5,7	0,6	4,0	13,2	3,3	0,7
10	0,3	79,7	14,0	13,0	92,9	12,7	2,1	0,7	3,3	1,1	5,9	1,0	0,7	2,6	0,6	6,0	19,6	3,3	0,6
11	0,3	87,2	16,0	13,0	81,3	4,5	1,1	1,1	2,6	0,0	5,7	1,4	1,3	3,6	0,7	4,0	12,7	3,2	0,4
12	0,3	82,6	16,0	15,0	93,8	1,4	0,3	0,3	0,8	0,0	6,3	1,6	1,3	3,8	0,6	4,0	13,0	3,3	0,7
13	0,3	77,4	16,0	16,0	100,0	0,1	0,0	0,0	0,0	0,0	7,2	1,8	1,8	4,9	0,6	4,0	13,0	3,2	0,7
14	0,4	86,3	17,0	16,0	94,1	5,2	1,3	1,2	2,7	0,0	3,2	0,8	0,2	1,2	0,7	4,0	13,2	3,3	0,7
15	0,3	85,7	16,0	16,0	100,0	0,1	0,0	0,0	0,0	0,0	6,7	1,7	1,2	3,7	0,6	4,0	13,1	3,3	0,6
16	0,3	88,1	16,0	16,0	100,0	1,0	0,3	0,4	1,0	0,0	5,4	1,4	1,1	3,3	0,6	4,0	13,1	3,3	0,7
17	0,4	86,9	16,0	16,0	100,0	0,8	0,2	0,3	0,6	0,0	6,7	1,7	1,2	3,5	0,6	4,0	13,0	3,3	0,7
18	0,3	76,5	16,0	16,0	100,0	0,7	0,2	0,3	0,6	0,0	4,7	1,2	0,8	2,6	0,6	4,0	12,9	3,2	0,7
19	0,2	80,6	16,0	16,0	100,0	1,3	0,3	0,5	1,3	0,0	4,4	1,1	0,6	2,1	0,6	4,0	13,2	3,3	0,7
20	0,3	86,0	16,0	16,0	100,0	5,9	1,5	1,3	3,0	0,0	4,0	1,0	0,6	2,0	0,6	4,0	13,2	3,3	0,6
21	0,3	78,6	16,0	16,0	100,0	0,8	0,2	0,3	0,8	0,0	4,2	1,1	0,7	2,3	0,6	4,0	13,1	3,3	0,6
22	0,3	84,7	16,0	16,0	100,0	5,9	1,5	0,8	2,4	0,1	6,7	1,7	1,8	4,8	0,6	4,0	13,1	3,3	0,7
23	0,3	83,9	16,0	16,0	100,0	0,1	0,0	0,0	0,0	0,0	7,0	1,8	1,1	3,2	0,6	4,0	13,1	3,3	0,6
24	0,4	76,4	16,0	16,0	100,0	3,3	0,8	0,6	1,8	0,0	3,6	0,9	0,3	1,5	0,7	4,0	13,0	3,3	0,6
25	0,3	87,4	16,0	16,0	100,0	3,5	0,9	1,5	3,4	0,0	9,0	2,2	1,8	5,0	0,7	4,0	13,1	3,3	0,7

Table A.13: Execution metrics - Part 6

N°	S4 M19	S4 M20	S4 M21	S4 M22	S4 M23	S4 M24	S4 M25	S4 M26	S4 M27	S4 M28	S4 M29	S4 M30	S4 M31	S4 M32	S5 M1	S5 M2	S5 M3	S5 M4	S5 M5
1	3,3	2,7	16,0	67,0	4,2	1,1	6,6	2,6	29,0	10,8	0,2	1,6	0,7	0,4	67,9	12,0	9,0	75,0	6,7
2	4,3	2,6	14,0	58,4	4,2	1,3	7,6	2,3	31,1	8,3	0,2	1,4	0,6	0,3	89,1	17,0	17,0	100,0	8,5
3	4,3	2,7	14,0	57,6	4,1	1,3	7,6	2,7	29,1	9,0	0,3	1,0	0,6	0,2	88,9	17,0	17,0	100,0	11,6
4	4,4	2,7	14,0	52,8	3,8	0,8	5,6	2,5	27,0	9,8	0,3	1,7	0,7	0,4	95,6	17,0	17,0	100,0	8,0
5	4,3	2,7	14,0	56,0	4,0	0,9	5,7	2,5	29,8	7,2	0,2	0,8	0,5	0,2	93,0	17,0	17,0	100,0	9,1
6	4,3	2,7	14,0	57,2	4,1	1,0	5,7	2,7	27,4	8,3	0,2	0,7	0,6	0,2	90,3	17,0	17,0	100,0	8,1
7	4,3	2,7	14,0	63,9	4,6	1,3	7,2	3,0	29,5	7,3	0,1	0,7	0,5	0,2	65,0	12,0	9,0	75,0	3,3
8	4,3	2,7	14,0	54,2	3,9	1,2	7,7	2,3	28,3	7,7	0,1	0,7	0,6	0,2	89,9	17,0	17,0	100,0	10,2
9	4,4	2,7	14,0	64,5	4,6	1,3	7,2	3,0	33,3	8,6	0,1	1,9	0,6	0,4	97,7	17,0	17,0	100,0	10,2
10	4,2	2,6	12,0	52,5	4,4	1,4	6,4	2,6	34,7	7,9	0,2	1,9	0,7	0,4	87,0	17,0	17,0	100,0	7,0
11	3,7	2,7	14,0	61,6	4,4	1,0	6,3	3,3	34,1	9,3	0,2	1,4	0,7	0,3	88,2	17,0	17,0	100,0	10,2
12	4,4	2,6	14,0	59,3	4,2	1,3	6,7	2,8	28,4	9,5	0,1	1,9	0,7	0,4	90,6	17,0	17,0	100,0	8,3
13	4,3	2,6	14,0	57,2	4,1	1,0	6,3	2,6	25,9	7,0	0,1	0,7	0,5	0,2	97,9	18,0	18,0	100,0	11,7
14	4,4	2,7	14,0	59,0	4,2	1,2	6,9	2,7	31,8	8,2	0,2	1,1	0,6	0,2	80,9	15,0	14,0	93,3	9,9
15	4,3	2,7	14,0	64,3	4,6	1,0	7,4	3,0	29,0	8,4	0,2	0,9	0,6	0,2	104,0	17,0	17,0	100,0	7,0
16	4,4	2,6	14,0	61,9	4,4	1,3	7,6	2,8	34,0	12,2	0,2	5,4	0,9	1,3	88,7	17,0	17,0	100,0	10,2
17	4,3	2,6	14,0	61,6	4,4	1,5	7,2	2,8	32,3	11,3	0,1	4,3	0,8	1,0	100,4	17,0	17,0	100,0	10,1
18	4,3	2,6	14,0	53,6	3,8	1,1	6,7	2,0	29,8	9,0	0,2	1,7	0,6	0,3	92,0	17,0	17,0	100,0	9,5
19	4,3	2,7	14,0	59,3	4,2	1,0	6,1	2,6	29,5	7,1	0,1	0,8	0,5	0,2	81,7	15,0	14,0	93,3	9,3
20	4,3	2,7	14,0	62,0	4,4	1,4	7,7	3,2	28,3	7,0	0,1	0,7	0,5	0,2	95,9	17,0	17,0	100,0	9,1
21	4,3	2,7	14,0	54,4	3,9	0,7	5,6	2,9	28,7	10,0	0,2	2,1	0,7	0,5	65,4	11,0	9,0	81,8	7,4
22	4,3	2,7	14,0	59,9	4,3	1,4	7,9	2,6	28,8	6,9	0,1	0,8	0,5	0,2	89,0	17,0	17,0	100,0	8,8
23	4,3	2,7	14,0	61,9	4,4	1,3	6,6	2,4	30,7	8,5	0,2	1,1	0,6	0,2	90,0	17,0	17,0	100,0	9,0
24	4,3	2,7	14,0	53,0	3,8	0,7	5,8	3,0	27,2	8,3	0,2	1,1	0,6	0,2	92,2	17,0	17,0	100,0	10,2
25	4,4	2,7	14,0	61,7	4,4	1,2	6,5	2,6	32,5	9,0	0,2	1,3	0,6	0,2	91,5	17,0	17,0	100,0	8,1

Table A.14: Execution metrics - Part 7

N°	S5 M6	S5 M7	S5 M8	S5 M9	S5 M10	S5 M11	S5 M12	S5 M13	S5 M14	S5 M15	S5 M16	S5 M17	S5 M18	S5 M19	S5 M20	S5 M21	S5 M22	S5 M23	S5 M24
1	0,7	0,5	2,0	0,1	19,1	1,9	1,4	5,1	0,6	10,0	32,7	3,3	0,6	4,3	2,6	10,0	40,1	4,0	1,0
2	0,5	0,2	1,1	0,4	7,6	1,9	1,2	3,8	0,6	4,0	12,6	3,2	0,6	4,1	2,6	16,0	61,6	3,8	0,7
3	0,7	0,6	2,4	0,4	7,7	1,9	1,4	4,3	0,7	4,0	12,8	3,2	0,6	4,2	2,6	16,0	61,8	3,9	0,8
4	0,5	0,5	2,4	0,0	9,2	2,3	1,4	4,3	0,6	4,0	12,9	3,2	0,7	4,2	2,6	16,0	69,3	4,3	1,4
5	0,5	0,6	3,0	0,0	8,7	2,2	0,9	3,0	0,7	4,0	12,9	3,2	0,6	4,2	2,6	16,0	68,6	4,3	1,1
6	0,5	0,2	1,2	0,4	8,6	2,2	1,8	5,1	0,7	4,0	12,9	3,2	0,6	4,2	2,6	16,0	66,6	4,2	1,5
7	0,3	0,7	2,5	0,0	19,0	1,9	1,1	3,6	0,7	10,0	32,8	3,3	0,6	4,4	2,6	10,0	39,7	4,0	0,9
8	0,6	0,8	3,8	0,0	6,7	1,7	0,7	2,4	0,7	4,0	12,9	3,2	0,7	4,3	2,6	16,0	63,0	3,9	0,8
9	0,6	0,5	2,5	0,2	8,4	2,1	1,2	4,1	0,7	4,0	12,8	3,2	0,6	4,2	2,6	16,0	71,2	4,4	1,4
10	0,4	0,3	1,5	0,0	10,2	2,6	2,0	5,1	0,6	4,0	12,7	3,2	0,6	4,1	2,6	16,0	62,6	3,9	1,0
11	0,6	0,6	3,1	0,1	8,0	2,0	1,7	4,9	0,7	4,0	12,8	3,2	0,6	4,2	2,6	16,0	61,7	3,9	1,0
12	0,5	0,6	2,9	0,0	6,8	1,7	0,8	2,6	0,7	4,0	12,9	3,2	0,7	4,2	2,6	16,0	66,8	4,2	1,1
13	0,7	0,8	3,2	0,0	6,3	1,6	1,2	3,5	0,7	4,0	12,8	3,2	0,6	4,2	2,6	16,0	66,3	4,1	1,2
14	0,7	0,8	2,9	0,0	11,3	1,9	1,0	3,2	0,6	6,0	19,1	3,2	0,5	4,2	2,7	14,0	52,3	3,7	0,9
15	0,4	0,5	2,4	0,0	10,1	2,5	1,7	5,2	0,7	4,0	12,9	3,2	0,6	4,2	2,7	16,0	75,1	4,7	1,8
16	0,6	0,8	3,6	0,0	7,4	1,8	1,5	4,4	0,6	4,0	12,7	3,2	0,7	4,2	2,6	16,0	63,0	3,9	1,3
17	0,7	1,2	5,2	0,0	9,3	2,3	1,2	3,7	0,7	4,0	12,8	3,2	0,6	4,2	2,6	16,0	64,8	4,1	1,4
18	0,6	0,7	3,2	0,0	9,8	2,4	1,7	5,2	0,6	4,0	12,9	3,2	0,6	4,2	2,6	16,0	66,7	4,2	1,2
19	0,6	0,7	2,5	0,0	15,2	2,5	1,9	6,2	0,7	6,0	19,1	3,2	0,5	4,2	2,7	14,0	54,7	3,9	1,3
20	0,5	0,3	1,4	0,1	9,5	2,4	1,2	3,9	0,7	4,0	13,0	3,2	0,6	4,2	2,7	16,0	70,0	4,4	1,3
21	0,8	1,1	3,1	0,0	16,6	2,1	1,4	5,4	0,6	8,0	26,0	3,3	0,6	4,3	2,6	10,0	39,2	3,9	1,0
22	0,5	0,5	2,3	0,0	10,2	2,6	1,4	4,4	0,7	4,0	12,9	3,2	0,6	4,3	2,6	16,0	63,6	4,0	0,9
23	0,5	0,8	3,8	0,0	6,8	1,7	1,1	3,1	0,7	4,0	12,8	3,2	0,6	4,2	2,6	16,0	65,2	4,1	1,3
24	0,6	0,4	1,7	0,4	7,8	1,9	2,0	5,3	0,6	4,0	12,9	3,2	0,7	4,3	2,6	16,0	65,6	4,1	1,2
25	0,5	0,4	1,6	0,0	5,1	1,3	0,5	1,9	0,6	4,0	12,8	3,2	0,7	4,2	2,6	16,0	68,6	4,3	1,9

Table A.15: Execution metrics - Part 8

N°	S5 M25	S5 M26	S5 M27	S5 M28	S5 M29	S5 M30	S5 M31	S5 M32	S6 M1	S6 M2	S6 M3	S6 M4	S6 M5	S6 M6	S6 M7	S6 M8	S6 M9	S6 M10	S6 M11	S6 M12
1	6,1	2,9	21,9	5,7	0,0	0,7	0,5	0,2	53,0	10,0	9,0	90,0	7,2	0,9	0,9	2,9	0,0	9,8	1,2	0,7
2	5,0	2,7	23,1	8,9	0,0	0,8	0,5	0,2	59,2	11,0	11,0	100,0	11,8	1,2	1,0	2,7	0,0	11,3	1,1	0,7
3	5,5	2,9	24,0	9,1	0,0	0,7	0,5	0,2	65,9	11,0	11,0	100,0	20,5	2,1	2,0	5,7	0,0	7,8	0,8	0,2
4	8,2	3,0	21,8	11,8	0,0	1,3	0,7	0,3	62,8	11,0	11,0	100,0	11,8	1,2	2,3	7,7	0,0	11,7	1,2	1,0
5	6,9	2,9	21,8	8,8	0,0	0,7	0,5	0,2	54,8	11,0	11,0	100,0	2,3	0,2	0,2	0,7	0,0	15,1	1,5	1,2
6	8,1	2,6	22,4	8,5	0,0	0,7	0,5	0,2	67,4	13,0	9,0	69,2	8,5	0,9	1,1	3,5	0,0	8,1	1,0	0,8
7	5,8	2,8	17,6	6,7	0,0	1,5	0,6	0,4	60,5	11,0	11,0	100,0	11,2	1,1	1,6	4,4	0,0	11,0	1,1	0,7
8	5,6	2,7	20,3	10,2	0,0	1,3	0,6	0,3	85,9	17,0	11,0	64,7	6,3	0,9	1,3	3,7	0,0	7,3	1,2	0,8
9	7,8	2,8	24,2	9,6	0,0	1,4	0,6	0,4	54,5	11,0	11,0	100,0	6,6	0,7	0,4	1,2	0,0	10,4	1,0	0,7
10	6,2	2,6	22,7	10,9	0,0	1,7	0,6	0,4	64,9	13,0	10,0	76,9	2,5	0,3	0,6	1,7	0,0	12,2	1,5	1,0
11	6,0	2,8	22,1	9,7	0,0	1,3	0,6	0,3	65,9	12,0	11,0	91,7	8,8	0,9	1,0	3,3	0,0	15,9	1,6	1,6
12	6,4	2,7	19,2	9,2	0,0	1,0	0,5	0,2	59,2	11,0	11,0	100,0	9,4	0,9	1,1	3,3	0,0	10,6	1,1	0,7
13	6,7	2,7	26,3	9,2	0,0	1,0	0,5	0,3	76,0	13,0	12,0	92,3	6,3	0,6	1,2	3,4	0,0	23,3	2,3	1,8
14	5,8	2,4	19,8	8,5	0,0	1,6	0,6	0,3	70,0	14,0	11,0	78,6	13,4	1,7	2,1	5,1	0,0	8,2	1,0	0,7
15	9,1	2,4	22,5	9,7	0,0	1,3	0,6	0,3	58,2	11,0	11,0	100,0	5,4	0,5	0,8	2,5	0,0	15,4	1,5	1,1
16	6,7	2,5	20,9	9,0	0,0	0,7	0,5	0,2	64,3	12,0	9,0	75,0	7,6	0,9	1,2	3,7	0,0	9,1	1,1	0,9
17	8,0	2,6	26,6	9,1	0,0	0,9	0,5	0,2	66,3	13,0	8,0	61,5	0,5	0,1	0,2	0,4	0,0	7,8	1,3	0,6
18	7,0	2,7	23,1	9,5	0,0	1,4	0,6	0,3	54,0	11,0	11,0	100,0	3,7	0,4	0,7	2,3	0,0	13,4	1,3	0,7
19	7,7	2,6	20,6	8,3	0,0	0,7	0,6	0,2	75,4	13,0	11,0	84,6	16,2	1,6	2,9	8,2	0,0	11,4	1,1	0,6
20	7,6	2,8	24,7	10,1	0,0	1,4	0,6	0,3	66,7	13,0	11,0	84,6	9,5	1,0	1,3	3,8	0,0	8,9	0,9	0,5
21	6,7	3,0	19,4	5,7	0,0	0,7	0,5	0,2	69,3	13,0	10,0	76,9	7,6	0,9	1,1	3,2	0,0	12,1	1,5	1,6
22	5,5	2,7	24,0	10,0	0,0	0,8	0,6	0,2	69,1	13,0	10,0	76,9	8,1	1,0	1,0	2,4	0,0	10,3	1,3	1,1
23	7,1	2,3	19,1	9,0	0,0	0,7	0,5	0,2	62,9	12,0	11,0	91,7	5,2	0,5	1,0	3,4	0,0	13,4	1,3	1,0
24	5,9	2,1	23,5	9,6	0,0	1,3	0,6	0,3	60,4	11,0	11,0	100,0	13,9	1,4	1,1	3,4	0,0	8,6	0,9	0,6
25	8,1	2,2	18,4	8,2	0,0	0,8	0,5	0,2	89,5	16,0	10,0	62,5	4,0	0,7	0,8	1,9	0,0	8,6	1,4	0,8

Table A.16: Execution metrics - Part 9

N°	S6 M13	S6 M14	S6 M15	S6 M16	S6 M17	S6 M18	S6 M19	S6 M20	S6 M21	S6 M22	S6 M23	S6 M24	S6 M25	S6 M26	S6 M27	S6 M28	S6 M29	S6 M30	S6 M31	S6 M32
1	2,6	0,6	8,0	26,3	3,3	0,6	4,3	2,6	8,0	32,3	4,0	1,7	6,7	1,5	53,0	6,8	0,2	3,4	0,9	1,0
2	2,8	0,6	10,0	32,7	3,3	0,5	4,3	2,6	8,0	34,1	4,3	1,2	6,0	2,8	59,2	4,9	0,4	0,9	0,6	0,2
3	1,3	0,6	10,0	33,2	3,3	0,5	4,3	2,6	8,0	32,2	4,0	0,7	4,9	3,2	65,9	4,9	0,2	0,7	0,6	0,2
4	4,2	0,6	10,0	35,1	3,5	0,6	4,6	2,9	8,0	32,5	4,1	1,9	7,7	0,5	62,8	6,6	0,2	2,9	0,8	0,8
5	4,7	0,6	10,0	33,2	3,3	0,5	4,3	2,7	8,0	30,6	3,8	1,3	6,8	2,2	54,8	5,0	0,4	0,7	0,6	0,1
6	3,2	0,6	8,0	26,8	3,3	0,6	4,3	2,6	10,0	41,3	4,1	1,1	6,5	2,9	67,4	6,1	0,1	1,2	0,6	0,3
7	3,0	0,6	10,0	32,8	3,3	0,5	4,3	2,6	8,0	30,6	3,8	0,9	5,7	2,8	60,5	5,4	0,6	0,8	0,7	0,1
8	2,9	0,6	6,0	19,8	3,3	0,5	4,3	2,6	14,0	53,5	3,8	1,1	6,5	2,7	85,9	8,1	0,2	1,0	0,6	0,2
9	3,1	0,6	10,0	33,1	3,3	0,5	4,3	2,6	8,0	31,7	4,0	1,1	6,5	2,8	54,5	4,3	0,3	0,7	0,5	0,1
10	3,8	0,6	8,0	26,8	3,4	0,6	4,3	2,6	10,0	39,5	3,9	1,0	6,2	2,9	64,9	5,8	0,1	0,7	0,6	0,2
11	5,7	0,6	10,0	33,0	3,3	0,5	4,3	2,6	8,0	33,4	4,2	1,2	7,0	3,2	65,9	5,4	0,2	1,6	0,7	0,4
12	2,7	0,6	10,0	33,0	3,3	0,5	4,3	2,6	8,0	34,1	4,3	1,3	6,7	2,5	59,2	4,0	0,1	0,7	0,5	0,2
13	5,8	0,6	10,0	32,6	3,3	0,4	4,1	2,6	10,0	47,2	4,7	1,6	8,3	2,6	76,0	7,0	0,5	1,2	0,7	0,2
14	2,8	0,6	8,0	27,0	3,4	0,6	4,3	2,6	10,0	37,7	3,8	1,3	7,4	2,9	70,0	6,7	0,6	0,8	0,7	0,0
15	3,8	0,6	10,0	33,1	3,3	0,5	4,3	2,6	8,0	32,3	4,0	1,3	7,1	2,7	58,2	4,8	0,2	0,7	0,6	0,2
16	3,4	0,6	8,0	26,9	3,4	0,6	4,2	2,6	10,0	40,1	4,0	1,0	6,5	3,0	64,3	5,9	0,1	1,0	0,6	0,3
17	2,4	0,7	6,0	20,0	3,3	0,7	4,3	2,6	10,0	43,8	4,4	1,7	8,1	3,0	66,3	5,3	0,2	1,0	0,5	0,2
18	2,7	0,6	10,0	33,2	3,3	0,5	4,3	2,6	8,0	31,2	3,9	0,5	4,8	3,1	54,0	4,5	0,2	0,7	0,6	0,2
19	2,2	0,6	10,0	33,4	3,3	0,5	4,3	2,6	10,0	39,1	3,9	0,9	5,7	2,8	75,4	4,9	0,1	0,7	0,5	0,2
20	1,9	0,6	10,0	33,6	3,4	0,5	4,4	2,7	8,0	32,4	4,0	1,2	6,9	2,8	66,7	3,4	0,2	0,7	0,4	0,2
21	5,3	0,6	8,0	26,9	3,4	0,6	4,3	2,6	10,0	39,4	3,9	1,2	6,8	2,5	69,3	5,7	0,2	1,6	0,6	0,4
22	3,7	0,6	8,0	27,0	3,4	0,6	4,3	2,6	10,0	38,9	3,9	0,7	5,2	3,2	69,1	5,4	0,2	0,7	0,5	0,2
23	3,8	0,6	10,0	32,9	3,3	0,5	4,3	2,6	8,0	33,5	4,2	1,1	6,2	3,0	62,9	4,6	0,3	0,7	0,6	0,2
24	2,6	0,6	10,0	33,3	3,3	0,5	4,3	2,7	8,0	33,6	4,2	0,9	6,2	3,2	60,4	4,8	0,4	0,7	0,6	0,1
25	2,8	0,6	6,0	19,2	3,2	0,5	4,1	2,6	14,0	56,9	4,1	0,9	6,1	3,0	89,5	7,3	0,2	0,7	0,5	0,2

Table A.17: Execution metrics - Part 10

ID	Comments
1	Quand robot défie notre logique perturbant et frustant. Globalement bien aimé. Notion de hierarchie, en fonction de la tâche (notamment simple) on se supérieur (on sait mieux faire) et donc le robot doit nous suivre. (2) bien pour tâche au + vite car va vite mais moins bien pour être lib, dépend de la tâche, pire (4)
2	RF est bien mais les mauvais choix sont pénible, frustrant, il ne concidère pas mon objectif/preferences. Cependant il s'adapte quand même une fois l'action faite. (2) RF trop cool, va + vite pour faire la tâche mais marche pas bien pour lib. HF est plus efficace en fonction de l'objectif. pire (4)
3	Chrono TO stressant. ne pas pouvoir prendre les cubes à l'avance n'est pas naturel, pertubant et rend un peu compliqué mais devient simple une fois habitué. Son de fin de tâche. Se sent obligé de faire quelque chose à chaque étape, et donc meme en RF clique sur la main pour confirmer que je serai passif. Rappeller à la fin de tache le regime et obj. (5) et (1) HF pour finir la tâche une fois habitué va relativement vite et fluide. HF pour etre libéré aussi, apres rien a faire.
4	Bon moment, Clair preference pour HF. RF est une catastrophe. En RF on n'a pas vraiment son mot à dire. En plus, quand RF commence à faire une erreur on est plus concentré à l'empecher de continuer à faire des erreurs plutot que sur la tâche, très frustant. Meme quand RF fait bon choix on se sent obligé de l'ecouter et impuissant.(5) HF et lib, car fluide en controle et j'ai pu atteindre mon objectif. Pire (6)
5	Intuitif, marche bien, efficace. Etre libere un peu frustant car on a envie d'agir, regarder le robot faire etre penible.. Sol: Donner une tache auxiliaire à l'humain a faire uniquement quand se désengage de la tache principale ? (3) HF et tache wrong. Car H agit le plus, R agit seulement quand necessaire. De plus, le robot s'adapte. Il a anticipé si jamais je ne prend pas le bleu en prenant le vert, mais une fois qu'il m'a vu prendre le bleu il n'a pas pris le sien. pire (4)
6	Globalement positif. Aurai aimé que le robot donne plus d'indication, guide plus les actions. (2) vif, mieux pour la tache, mais moins prévisible. (1) bien aussi mais plus lent/passif. pire le (6)
7	Certain scenario vraiment contraignant et frustant. Uiliser ressource commune en 1er est vraiment un mauvaix choix. Globalement ok. 2 efficace une fois compris. Faire la dernière action est gratifiant, donc que le robot s'adapte pour le permettre s'est positif.Preféré (2), rapide et bon choix, pire 4

Table A.18: Comments from participants given after the experiment. Part 1

ID	Comments
8	Parfois frustant a cause des mauvais choix du robot. Les actions sont "rigide", pas naturel (attraper cube bleu sous vert, le faisant "tomber" le vert). Bien dans l'ensemble, le fait que chacun ai sa part rend la collaboration pertinente et utile. Je prefère que la priorité des choix et actions soit à l'humain. (3) car s'adapte à ce que je fais et c'est clair. Pire (4) un mauvais choix du robot a impliqué un mauvais choix de ma part, frustrant (1) simplement frustrant, le robot semble contre moi.
9	RF pas efficace, mais follow est plus simple, moins compliqué. HF mieux mais parfois incohérent (barre rose). (1) est le plus satisfaisant. (4) est le pire
10	ne pas pouvoir attraper en avance un peu agaçant. Le fait de devoir regarder le but à gauche + la scène + lire le prompt text un peu complex =>mieux si robot dit quoi faire.Prefère quand le robot est "passif", dans le sens follower. Que le robot attende qu'on decide puis agisses. N'aime pas quand le robot "prend des initiatives" car possible mauvais choix: vole les cubes très agaçant.. + prend cube commun en 1er. Pref 1 Pire 4
11	Bonne simu, claire. Mvt ont l'air réel tache claire et globalement se passe bien. Les steps cadense bien, pratique Meilleur (1) pire (4)
12	Temps d'attente (step) frustrant. Serait bien de pouvoir prendre avant pour indiquer intention, donner info. HF + interessant + de control, - efficace mais - frustrant, - imprevu et donc de mauvais choix.(+) 1 (-) 6 car on est obligé de reposer le cube.
13	Interaction simple, comme un jeu video. Les mauvais choix du robot sont assez frustrant. Simple et plaisant.(+) 3 car s'est bien adapté malgré erreur humaine (-) 4 car vole les cubes
14	Tres bien dans l'ensemble. Géné par l'affichage, du mal à lire. Certain scenario efficace d'autre non. Simple, sauf lecture/texte. (+) 5 fini rapidement, double satisfaction de finir sa part vitre puis voir la pile fini par le robot (-) 6 degouté, frustrant
15	Bien, a volé 2 fois les cubes, pas agréable. HF + facile.(+) 2 (-) 4, 3
16	Simple, agréable. Le manque de collision avec cube réduit le réalisme mais ok. Un peu confu/pertubant et un peu frustrant de devoir attendre que robot finisse action avant d'agir à nouveau.(+) 2, RF car rapide (-) 6

Table A.19: Comments from participants given after the experiment. Part 2

ID	Comments
17	certain scenario ok =>c'est plutot H qui a mal agit, 2 sce avec mauvais choix de R. HF/RF pas tellement choix !=(-) 6 RF etre lib (+) 2
18	pas pouvoir prendre cube en avance frustrant. En RF, R devient prévisiblement génant, on prévoit et réfléchi pour s'adapter et anticiper mauvais choix (defensif). HF mieux.(+) HF tache plus vite, plus interessant que lib rapidement, ennuyeux. (-) 4, R imprévisible, devient ennemi
19	Globalement ok reagit bien, comprend bien ce que je faisait (pink), avant dernier tres frustrant, vole les cubes.. Sinon bien passé (+) 2 (-) 4
20	Simu claire, voit bien le R et quand il agit, interaction bonne et claire, 4 mauvais mais globalement benefique interaction. Pref au quotidien laisser R faire les tache chronophage, donc bien aimé obj lib au plus vite.(+) 5 puis 2 (-) 4
21	R prévisible cool, on peut preshot et anticiper. (+) 1 3 (-) 4
22	Sympa, simu bien faite. interactif, on est pris dedans et acteur. Beaucoup d'émotion déjà (satisfaction / frustration) donc si c'était une tache plus concrete ça serait encore plus frustrant.(+) 1 3 (-) 4
23	Simu bien faite, s'imagine bien interaction. Au debut un peu confu diff entre HF/RF puis ok. Confusion cube blanc et gris(+) aucun (-) 6
24	Interessant, jamais une gène, interaction amusante, sympa de prédire RA(+) 1, 3 (-) 6
25	HF ok, qd RF on peut rien faire. Simulation moins naturelle, pas parfaitement réaliste. Notion etape/synchronisation perturbant un peu (+) 5 HF lib (-) 4

Table A.20: Comments from participants given after the experiment. Part 3

	S1	S2	S3	S4	S5	S6
Times preferred the most	***9***	***7***	***3***	*0*	***5***	*0*
Times preferred the least	*0*	*0*	*0*	***16***	*0*	***9***

Table A.21: Number of times each scenario has been respectively preferred the most and the least. HF scenarios are never disliked and RF scenarios with erroneous estimations are never preferred.

Bibliography

[Cohen 1988] Jacob Cohen. *The concepts of power analysis. Statistical power analysis for the behavioral sciences.* Hillsdale: Elrbaum, 1988. (Cited in page 11.)

[Devin 2018] Sandra Devin, Camille Vrignaud, Kathleen Belhassein, Aurelie Clodic, Ophelie Carreras and Rachid Alami. *Evaluating the Pertinence of Robot Decisions in a Human-Robot Joint Action Context: The PeRDITA Questionnaire.* In 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), pages 144–151, Nanjing, August 2018. IEEE. (Cited in pages 2 and 3.)