Designing Interaction for Human-Machine Collaboration in Multi-Agent Scheduling

by

Jorge Perez

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science
at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2015

© Massachusetts Institute of Technology 2015. All rights reserved.

Author .................................................................
Department of Electrical Engineering and Computer Science
August 18, 2015

Certified by ...........................................................
Prof. Julie Shah
Associate Professor
Thesis Supervisor

Accepted by ..........................................................
Prof. Christopher Therman
Chairman, Department Committee on Graduate Theses
Designing Interaction for Human-Machine Collaboration in
Multi-Agent Scheduling

by

Jorge Perez

Submitted to the Department of Electrical Engineering and Computer Science
on August 18, 2015, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

In the field of multi-agent task scheduling, there are many algorithms that are capable
of minimizing objective functions when the user is able to specify them. However,
there is a need for systems and algorithms that are able to include user preferences
or domain knowledge into the final solution. This will increase the usability of al-
gorithms that would otherwise not include some characteristics desired by the end
user but are highly optimal mathematically. We hypothesize that allowing subjects
to iterate over solutions while adding allocation and temporal constraints would allow
them to take advantage of the computational power to solve the temporal problem
while including their preferences. No statistically significant results were found that
supported that such algorithm is preferred over manually solving the problem among
the participants. However, there are trends that support the hypothesis. We found
statistically significant evidence (p=0.0027), that subjects reported higher workload
when working with Manual Mode and Modification Mode rather than Iteration Mode
and Feedback Iteration Mode. We propose changes to the system that can provide
guidance for future design of interaction for scheduling problems.

Thesis Supervisor: Prof. Julie Shah
Title: Associate Professor
Acknowledgments

First, I would like to express my gratitude for Professor Julie Shah who has supported and guided me through this process. Interactive Robotics Group has been my home for over 2 years and Julie has been instrumental in my development as a researcher and engineer.

Furthermore, I am grateful to Vaibhav Unhelkar who first welcomed me into the group as my undergraduate research advisor and was a great mentor throughout my time at IRG. I would also like to thank Matthew Gombolay for his guidance and motivation when I doubted myself. I also thank my friends who helped keep me motivated and happy.

Finally, I would like to thank my parents, my sister and my girlfriend who have supported me through the ups and downs of MIT and this Thesis.
# Contents

1 Introduction ........................................... 13

2 Background ........................................... 15

3 Aim of Experiments ................................ 19
   3.1 Independent Variables ............................. 19
   3.2 Hypotheses ........................................ 20

4 Experimental Methods ................................ 23
   4.1 Graphical User Interface ......................... 23
      4.1.1 Allocation .................................... 24
      4.1.2 Sequencing ................................... 25
      4.1.3 Temporal Constraints ......................... 25
      4.1.4 Scores ......................................... 25
   4.2 Problem Design ...................................... 25
   4.3 Experimental Design ................................. 27
   4.4 Procedure .......................................... 28
   4.5 Objective Evaluation ............................... 29
   4.6 Subjective Evaluation .............................. 29

5 Results .................................................. 33
   5.1 Participants ......................................... 33
   5.2 Subjective Measures ............................... 34
   5.3 Objective Measures ................................. 35
6 Discussion

6.1 Reducing Workload ........................................ 37

6.2 Producing Objectively Better Solutions ................... 38

6.3 Gaining the Trust of the Users ............................... 38

6.4 Including User Preferences ................................. 38

7 Conclusion .................................................. 41

A Tables ...................................................... 43

B Figures ...................................................... 51
List of Figures

4-1 The graphical user interface. ........................................ 23
4-2 The allocation panel. ............................................. 24
4-3 The sequencing panel. ............................................ 24
4-4 The temporal constraints panel. ................................. 25
4-5 Cumulative Ergonomic Score vs Makespan for Build Car. .. 26
4-6 Cumulative Ergonomic Score vs Makespan for Build Desk. .. 26
4-7 Cumulative Ergonomic Score vs Makespan for Build House. .. 27
4-8 Cumulative Ergonomic Score vs Makespan for Build Airplane. .. 27

B-1 Answers to question 7 after using Manual mode. Median: 2 .... 51
B-2 Answers to question 7 after using Modification mode. Median: 2.5 . . 51
B-3 Answers to question 7 after using Iteration mode. Median: 4 .... 52
B-4 Answers to question 7 after using Feedback Iteration mode. Median: 4 . . 52
B-5 Answers to question 7 after using Manual mode. Median: 2 .... 52
B-6 Answers to question 7 after using Modification mode. Median: 2.5 . . 53
B-7 Answers to question 7 after using Iteration mode. Median: 4 .... 53
B-8 Answers to question 7 after using Feedback Iteration mode. Median: 4 . . 53
B-9 Total Score for Manual Mode .................................... 54
B-10 Total Score for Modification Mode ............................... 54
B-11 Total Score for Iteration Mode ................................. 54
B-12 Total Score for Feedback Iteration Mode ........................ 55
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Invalid Solutions per Mode</td>
<td>29</td>
</tr>
<tr>
<td>4.2</td>
<td>Post-Scenario Questionnaire</td>
<td>31</td>
</tr>
<tr>
<td>4.3</td>
<td>Final Questionnaire</td>
<td>32</td>
</tr>
<tr>
<td>5.1</td>
<td>P values resulting from running Friedman Test</td>
<td>33</td>
</tr>
<tr>
<td>A.1</td>
<td>Build Car Agent Capabilities and Ergonomic Scores</td>
<td>44</td>
</tr>
<tr>
<td>A.2</td>
<td>Build Car Delays</td>
<td>44</td>
</tr>
<tr>
<td>A.3</td>
<td>Build Car Deadlines</td>
<td>45</td>
</tr>
<tr>
<td>A.4</td>
<td>Build Desk Agent Capabilities and Ergonomic Scores</td>
<td>45</td>
</tr>
<tr>
<td>A.5</td>
<td>Build Desk Delays</td>
<td>45</td>
</tr>
<tr>
<td>A.6</td>
<td>Build Desk Deadlines</td>
<td>46</td>
</tr>
<tr>
<td>A.7</td>
<td>Build House Agent Capabilities and Ergonomic Scores</td>
<td>46</td>
</tr>
<tr>
<td>A.8</td>
<td>Build House Delays</td>
<td>46</td>
</tr>
<tr>
<td>A.9</td>
<td>Build House Deadlines</td>
<td>47</td>
</tr>
<tr>
<td>A.10</td>
<td>Build Airplane Agent Capabilities and Ergonomic Scores</td>
<td>47</td>
</tr>
<tr>
<td>A.11</td>
<td>Build Airplane Delays</td>
<td>47</td>
</tr>
<tr>
<td>A.12</td>
<td>Build Airplane Deadlines</td>
<td>48</td>
</tr>
<tr>
<td>A.13</td>
<td>Median Scores for Post-Questionnaire</td>
<td>48</td>
</tr>
<tr>
<td>A.14</td>
<td>Median Objective Measures</td>
<td>48</td>
</tr>
<tr>
<td>A.15</td>
<td>P values resulting from running Wilcoxon Signed-Rank Test on Subjective Measures</td>
<td>49</td>
</tr>
<tr>
<td>A.16</td>
<td>P-Values resulting from running Repeated Measures Anova on Objective Measures</td>
<td>49</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

There is growing presence of automation to supplement lean practices in manufacturing lines. The capabilities of the agents in a team are different. In order to best allocate the resources, there is a need for algorithms that optimize the allocation and sequencing of tasks taking into account the capabilities of the agents, temporal constraints and spatial constraints.

State-of-the-art task and scheduling algorithms are able to efficiently generate near optimal schedules for multi-agent problems with temporal and spatial-proximity constraints [8]. However, these algorithms have several limitations implicit in their formulation, which can limit their usefulness in practical applications for human teams.

Scheduling algorithms can naturally handle objective functions that minimize makespan, idle time and other measures than can be expressed mathematically. However, end users may have trouble expressing the preferences or the intuition behind their choices. Thus, it is difficult to encode their intuition in the regular framework of scheduling algorithms. Moreover, the lack of apparent transparency of the inner workings of the algorithms lead to a lack of trust from the end user. The end user most of the time will have limited knowledge as to the criteria the algorithm is using to allocate and sequence tasks or about how the algorithm handles corner cases. If end users do not trust the algorithm, they might default to their own solution. Most of the time the user’s solution will be worse in terms of makespan [8] than the solution of an algorithm, but it will satisfy the user’s preferences.
In this work, we study the design of a system that allows users to include their preferences in the schedule without the need to explicitly include their preferences in the problem formulation. We seek to design a system that harnesses the computational power of computers and the environment awareness of humans to produce solutions that are mathematically near-optimal that are well informed with regards with the real world.

In this experiment, we test the effectiveness of four different modes. Manual mode in which the subjects manually solves a scheduling problem using a graphical user interface. Modification Mode is the same interaction as the manual mode but the subject starts with a valid solution produced by Tercio [8]. Iteration Mode in which the subjects starts with a valid solution from Tercio but is able to iterate over Tercio solutions while adding constraints. Feedback Iteration Mode where the interaction is the same as Iteration Mode but the graphical user interface gives feedback on the changes made after every iteration.

We hypothesize that subjects will achieve better overall performance and report less workload when using the Iteration Mode and Feedback Iteration Mode. Moreover, we hypothesize that subjects will prefer and trust more the Iteration Mode and Feedback Iteration Mode rather than Manual Mode or Modification Mode. A 16 participant human subject experiment was conducted. In the experiment, the subjects used all four different systems to solve scheduling problems. We found statistically significant evidence (p=0.0027), that subjects reported higher workload when working with Manual Mode and Modification Mode. Furthermore, we found trends that support that better overall performance can be achieved when using Iteration Mode and Feedback Iteration Mode.
Chapter 2

Background

Multi-agent scheduling has been the focus of research in many fields in order to increase the efficiency of their operations. The algorithms developed for these tasks are based on a mathematical formulation that is most of the time hidden from the end user, [8], [3]. The intrinsic limitation of these algorithms is the lack of power of the users to include their preferences on the fly. Their preferences would have to be translated into a mathematical function, which the algorithm can minimize. However, users do not always know how to express their preferences or how to quantify them.

In order to address these issues, there have been efforts to include humans in the decision making loops to enable the domain-specific knowledge to influence the final solution [4], [6], [5], [2].

In [15], the authors explored the problem of resource allocation with temporal constraints in the aircraft carrier deck. The planning is currently done without the help of any automation decision aids. The authors developed an integer-based linear program-based planning algorithm. They allowed the human to set both the goals and the constraints for the algorithm. The algorithm would then return the proposed solution for approval. The authors found that the expert humans were able to outperform the optimization algorithm by using their own heuristics. Thus, computerized solutions by themselves will not always outperform human experts in the computational aspects of problems. This is especially true for task scheduling problems for which the optimal solution cannot be calculated in polynomial time [1]. When using
pure automation, it is not trivial to include these heuristics in the algorithm since the experts cannot always explain their heuristic. We hypothesize that designing a system where the human and computer iterate to come up with a solution will allow for better inclusion of user heuristics and preferences.

The trust of humans in the system is critical to the success of the system. Research efforts to model the effects of trust in automation include [11], [7], [13], [10]. In [7], participants initially trusted the automation. However, they lost trust in the system after observing mistakes, even if the system was reliable, unless an explanation of the automation was provided. Scheduling systems can many times output solutions, which seem to be erroneous to the user but are in fact in accordance with the input problem formulation. Since the end user might not be familiar with the inner workings of the algorithm, a solution that does not take into account some piece of domain knowledge critical to the scenario might result in the user losing trust in the system. We hypothesize that allowing the end users to include their preferences might mitigate this effect.

The authors in [11] assert that the choice of manual or automatic control is an important factor that can affect performance of the system. They developed a model that defines human trust in machines, which helps researchers understand how trust changes the experience with a system.

Recent work in [10], shows that the issue of trust in automation is still important when designing current autonomous systems. The authors go further to compare the trust in automation to interpersonal trust, and establish that trust in automation, like interpersonal trust, can be hard to regain once it is lost. They conclude that absolute trust is unobtainable. However, facilitating the exploration of trust should help the end users achieve their task in a more effective manner.

Other related work focuses on studying the different types and levels of automation [14]. In this work, the authors developed a framework for appropriately selecting types and levels of automation that aim to collaborate with the human rather than replace them. They proposed four classes of functions for which automation can be applied: information acquisition, information analysis, decision and action selection,
and action implementation. The system used in this work helps the process of information acquisition by providing a visual representation of the agent’s capabilities and ergonomic scores. Our system also aids in the process of information analysis by providing aid in visualizing the temporal constraint and in the process of decision selection through the help of Tercio [8], a fast scheduling algorithm with temporospatial constraints.

In [9], the authors develop an interactive scheduling system in which the computer automatically produces a Gantt Chart using some heuristic, which ignored all subjective factors. The user was then able to make amends to the Gantt chart to include subjective factors. In our experiments, we model one of our independent variables (Modification Mode) after this model. In this mode, the user will start with a valid solution from Tercio and then edit it to include his or her preferences.

A related effort, [12], studied the human role in scheduling in environments where the events change the specification of the problem and the scheduling of the tasks must be done in a group. In this case, decision making is uncertain and contextual information needs to be included in the schedule. The authors claim that human collaborative scheduling in these situations is central to achieving a good solution.

In this work, we aim to study intuitive interactions that allow humans to include their preferences or domain knowledge in the final solution. We are motivated by task scheduling problems that involve time sensitive or critical decisions, which require a reliable system that allows users to include their preferences.
Chapter 3

Aim of Experiments

In the experiments, we seek to understand the effectiveness of different systems in producing solutions with short makespans but include the preferences of the users. We provided users with ergonomic scores, which served as a proxy to user preferences or domain specific knowledge not embedded into the algorithm. Ergonomic scores in this case represent the physical strain a certain task will put on the agent’s body. Subjects were told that the algorithm did not take into account ergonomic scores when calculating the solution and were asked to minimize the total score, which is the sum of the makespan and the cumulative ergonomic score. The goal is to enable subjects to interact with different systems to solve scheduling problems in a scenario where humans have knowledge outside of the scope of the algorithm. This will allow us to better understand which features enable an interaction where the users have enough decision power to include their preferences, while taking advantage of the computational powers of computers.

3.1 Independent Variables

In this experiment, we control the level of assistance that is provided to the subject. There are four conditions in this experiment:

1. Manual Mode- The subject allocates tasks to agents and assign start times to the tasks manually in a manner that maximizes the score.
2. Modification Mode- The subject is given the output of one run of the Tercio scheduling algorithm as a starting point. They are asked to modify the existing solution to increase the score.

3. Iteration Mode- The subject is given the output of one run of the Tercio scheduling algorithm as a starting point. The subjects iteratively improve the solution by adding allocation and temporal constraints and running Tercio.

4. Feedback Iteration Mode- The subject is given the output of one run of Tercio as a starting point. The subjects iteratively improve the solution by adding allocation and temporal constraints and running the feedback modification algorithm. After each iteration, the algorithm provides feedback in the form of text.

3.2 Hypotheses

H1 Subjects prefer using Feedback Iteration Mode and Iteration Mode in the decision making process rather than using Modification Mode or Manual Mode. A series of subjective Likert-scale questions will be used to determine which modes subjects preferred.

In Manual Mode and Modification Mode, the subjects will need to familiarize themselves with the temporal constraints and form a temporally consistent solution. As opposed to the Manual Mode and Modification Mode, the Feedback Iteration Mode and Iteration Mode automatically handle the temporal optimization for the subjects. Thus, we expect that the reported workload will be much higher for the first two modes.

Moreover, Manual Mode and Modification Mode give more freedom to the users by giving the power to directly control the allocation and sequencing completely. However, we expect the increased workload from the temporal problems to overshadow the benefits from the first two modes.
H2 Schedules resulting from the interaction from a subject with Feedback Iteration Mode and Iteration Mode will be objectively better according to sum of the total time to complete the process and the cumulative ergonomic score.

Minimizing the makespan of a task scheduling problem with hard upper-bound and lower-bound temporal constraints is an NP-Hard problem [1]. However, recent algorithms [8] are capable of achieving experimentally near optimal schedules in terms of makespan. Given the heavy computational nature of the problem, we expect the humans subjects to perform worse than the algorithm, in terms of makespan.

In Modification Mode and Manual Mode, we expect humans to perform objectively slightly better than the Feedback Iteration Mode and the Iteration Mode, in terms of ergonomic score. Nevertheless, we expect that the difference in the inclusion of the ergonomic score will not be enough to compensate for the loss of optimality, in terms of time.

H3 Subjects will be more satisfied with the solutions resulting from Feedback Iteration Mode or Iteration Mode than the solutions produced by Modification Mode or Manual Mode.

This will be assessed through two Likert-scale questions, questions 7 and 10 in the post-scenario questionnaire in Table 2.

H4 The reported workload will be higher for the Manual Mode and Modification Mode than for the Feedback Iteration Mode and Iteration Mode.

We expect that the work done by the algorithm will lessen the workload for the Iteration Mode and Feedback Iteration Mode. In those modes, the users do not have to produce temporally consistent solutions manually, which we think is an important factor, in terms of workload. Six Likert-scale questions to assess workload will be administered as part of the post-scenario questionnaire.
H5 *Subjects will report a higher level of trust to consistently produce solutions for Feedback Iteration Mode or Iteration Mode than for Modification Mode or Manual Mode.*

In Manual Mode and Modification Mode, the final solution is always produced by the user. We expect subjects to be unsure about their performance in the makespan optimization portion of the total score. We hypothesize this will result in a lack of trust for the final solutions produced by these two modes. A subjective Likert-scale question, question 8 in Table 2, will be used to assess the reported level of trust.
Chapter 4

Experimental Methods

The experiment consisted of solving four different task scheduling problems using the four different systems, which were previously presented. Each subject solved all four problems and used all four modes. Each problem consisted of 14-16 tasks, 12-16 lower-bound temporal constraints, 2-3 upper-bound constraints, 2 human agents and 1 robot agent. The goal of the subject in every mode was to use the system to produce a valid solution that minimized the total score. The total score was the simple sum of the total time to complete all the tasks and the cumulative ergonomic scores.

Figure 4-1: The graphical user interface.

4.1 Graphical User Interface

The subjects used a graphical user interface (Fig 1) to solve the scheduling problems. The graphical user interface consisted of three major components: an allocation panel
(Fig 2), a sequencing panel (Fig 3), and a temporal constraints panel (Fig 4).

Figure 4-2: The allocation panel.

4.1.1 Allocation

The allocation panel (Fig 2) provided the information about the agent capabilities and the ergonomic scores for each task. In Manual Mode and Modification Mode, the subjects used the allocation panel in Fig 2 to perform task allocation by clicking in the appropriate boxes. The task would then appear in the sequencing panel (Fig 3) in the first spot where it would not overlap with other tasks. In Feedback Iteration Mode and Iteration Mode, the allocation was performed by Tercio but the subjects could use the allocation panel to disallow any agent-task pairing.

Figure 4-3: The sequencing panel.
4.1.2 Sequencing

The sequencing panel (Fig 3) was the same for all of the modes. The subjects were able to drag the tasks along the timeline to assign them a start time. If a pair of tasks overlapped or violated a temporal constraint, both of them would automatically turn red. The sequencing panel also included a visual representation of the cumulative ergonomic score, shown as a blue bar at the bottom of the panel.

![Figure 4-4: The temporal constraints panel.](image)

4.1.3 Temporal Constraints

The temporal constraints panel (Fig 4) was used to display the temporal constraints. In the case of the Feedback Iteration Mode and the Iteration Mode, the temporal constraints panel could also be used to add temporal constraints using a drop-down menu as shown in Fig 4.

4.1.4 Scores

The total time duration, the cumulative ergonomic score, and the total score were all shown at the bottom of the allocation panel and updated automatically after any change to the allocation or to the start times.

4.2 Problem Design

Four problems were used throughout the experiment. Each subject solved every problem, albeit with a different mode. We sought to design four problems of similar difficulty which were easy enough to be solved by the subjects manually. However,
the problems were to be difficult enough, such that the subject could not trivially optimize for time and for ergonomics. Moreover, the four problems were designed such that the effects of ergonomics and time were similar. The four problems are simplifications of building processes.

1. Build Car: refer to Table 5, 6

2. Build Desk: refer to Table 7, 8

3. Build House: refer to Table 9, 10

4. Build Airplane: refer Table 11, 12

Figures 5-8 show the plots of the ergonomics and makespans scores of the valid solutions produced by running Tercio for 150 iterations on the four problems. The plots show that there were many possible valid solutions and possible scores. Note that
Tercio has no guarantee of optimality of makespan and does not take into account the ergonomic scores by design. Thus, some of the subjects achieved a total score lower than the first 150 iterations of Tercio.

4.3 Experimental Design

The aim of the experiment was to study the effectiveness of the different systems in optimizing for makespan, while including the preferences of the users. The experiment was designed to compare the systems among participants. However, the graphical user interface used for the different systems was very similar, so learning effects could play a factor in the results. In order to mitigate the learning effects, we randomized the order in which the subjects used the systems. Moreover, the four problems presented in section 4.2 are similar but they have some key differences that could influence the results. To block the effects of the different problems, we randomized the systems-
problem pairings.

4.4 Procedure

The subject was first given an overview of the elements of the graphical user interface. We explained where to find the tasks that needed to be performed, the two human agents and robotic agent, the agents’ capabilities, the ergonomic scores, the temporal constraints and the cumulative scores. The temporal constraints were explained in detail and two sheets, one explaining delays and one explaining deadlines, were provided for the subject to consult throughout the experiment.

Then, we proceeded to explain the steps to solve a task scheduling problem using the interface. Moreover, subjects were told that to solve a task scheduling problem they must allocate each task to an agent and assign a start time to each task. Subjects were told that the graphical user interface allowed them to check if the solution was valid by coloring in red the tasks that violated a constraint. The goal of minimizing the sum of the total time to complete the tasks and the cumulative ergonomic score was explained. Towards the end of the initial presentation, we introduced all four systems, but subjects were told that the specifics for each problem would be explained before each mode.

Before each mode, the subjects received training on how to use the interface to solve the problems. Moreover, the subjects were told that Tercio does not take into account ergonomic scores when calculating the solutions. Subjects were able to save solutions during the trial and the best solution was recorded as the final score for that trial. Finally, the subjects were given a pen and paper and were told they had 8 minutes to complete the task. Each trial consisted of solving a task scheduling problem using the graphical user interface. After each mode, the subjects were asked to respond to a post-scenario questionnaire (Table 2). At the end of the experiment, the subjects were asked to complete a final questionnaire (Table 3).
4.5 Objective Evaluation

Objective measures to rate the quality of the final solution include the makespan of the final solution, the cumulative ergonomic score and the total score.

1. **Makespan:** the time at which the last task is completed in the final solution provided by the user.

2. **Ergonomic Scores:** Ergonomic scores represented the relative strain that performing a certain task would put on the agents performing it. Subjects were provided with an ergonomic score on the scale 0-100, higher being more strain on the human, for each agent-task pairing. The robotic agent always had an ergonomic score of zero. We recorded the sum of all the ergonomic scores for each allocation in the final solution.

3. **Total Score:** the total score was the simple sum of the makespan and the cumulative ergonomic score.

<table>
<thead>
<tr>
<th>Table 4.1: Invalid Solutions per Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>Valid</td>
</tr>
<tr>
<td>Invalid</td>
</tr>
</tbody>
</table>

In some cases, the subjects were not able to provide a solution that satisfied all of the temporal constraints. The occurrence per mode is shown in Table 1. We recorded the validity of each solution and did not use the scores of invalid solutions in the statistical analysis.

4.6 Subjective Evaluation

After each problem, subjects were asked to fill a post-scenario questionnaire, which consisted of ten Likert-Scale question, as shown in Table 2. All four questions aim to understand how the subject rated different dimensions of the system’s effectiveness.
After using all four systems, the subjects were asked to complete a Final Questionnaire, which consisted of 6 Likert-scale question comparing the 6 possible pairings among the 4 systems.
Table 4.2: Post-Scenario Questionnaire

Please rate the following statements on a scale on 1 to 5.

<table>
<thead>
<tr>
<th>Cognitive Load:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mental Demand:</td>
</tr>
<tr>
<td>1. How mentally demanding was the task?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Physical Demand:</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. How physically demanding was the task?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Temporal Demand:</th>
</tr>
</thead>
<tbody>
<tr>
<td>3. How hurried or rushed was the pace of the task?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Performance:</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. How successful do you think you were in accomplishing what you were asked to do?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effort:</th>
</tr>
</thead>
<tbody>
<tr>
<td>5. How hard did you have to work to accomplish your level of performance?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frustration;</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. How irritated, stressed, annoyed, or discouraged were you?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Algorithm Evaluation:</th>
</tr>
</thead>
<tbody>
<tr>
<td>7. I am satisfied with the resulting schedule.</td>
</tr>
</tbody>
</table>

| 8. I trust the system to consistently produce schedules with which I am satisfied. |

| 9. The system allowed me to incorporate my preferences. |

| 10. I have a good understanding of why the system chose the schedule it did. |
Table 4.3: Final Questionnaire

Please rate the following statements on a scale on 1 to 5.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>In order to solve scheduling problems effectively,</td>
<td></td>
</tr>
<tr>
<td>I prefer using Manual Mode rather than using Iteration Mode.</td>
<td></td>
</tr>
<tr>
<td>I prefer using Manual Mode rather than using Feedback Iteration Mode.</td>
<td></td>
</tr>
<tr>
<td>I prefer using Modification Mode rather than using Iteration Mode.</td>
<td></td>
</tr>
<tr>
<td>I prefer using Modification Mode rather than using Feedback Iteration Mode</td>
<td></td>
</tr>
<tr>
<td>I prefer using Iteration Mode rather than using Feedback Iteration Mode.</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 5

Results

In this section, we present the results of running statistical tests on subjective and objective measures. We define the significance level as $\alpha = 0.05$ for the Omnibus Friedman tests and include the Bonferroni correction $\frac{\alpha}{6}$ for pairwise tests.

5.1 Participants

A total of sixteen participants completed the experiment. Each participant experienced all four modes, albeit in different order to control for learning effects. There were 6 women and 10 men among the participants with an average age of 23.1 ± 2.0.

<p>| Table 5.1: P values resulting from running Friedman Test |</p>
<table>
<thead>
<tr>
<th>---------------------------------</th>
<th>-----------</th>
<th>-----------</th>
<th>-----------</th>
<th>-----------</th>
<th>-----------</th>
<th>-----------</th>
<th>-----------</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0076</td>
<td>0.2059</td>
<td>0.0209</td>
<td>0.0027</td>
<td>0.0833</td>
<td>0.0956</td>
<td>0.7055</td>
</tr>
<tr>
<td>2</td>
<td>0.2080</td>
<td>0.0833</td>
<td>0.3173</td>
<td>0.5637</td>
<td>0.1573</td>
<td>0.0833</td>
<td>1.0000</td>
</tr>
<tr>
<td>3</td>
<td>0.0089</td>
<td>0.0114</td>
<td>0.0578</td>
<td>0.0075</td>
<td>0.5271</td>
<td>0.5271</td>
<td>0.1573</td>
</tr>
<tr>
<td>4</td>
<td>0.3017</td>
<td>0.3657</td>
<td>0.5930</td>
<td>0.0522</td>
<td>0.7630</td>
<td>0.3657</td>
<td>0.3173</td>
</tr>
<tr>
<td>5</td>
<td>0.2147</td>
<td>1.0000</td>
<td>0.4795</td>
<td>0.0578</td>
<td>0.3657</td>
<td>0.0956</td>
<td>0.3657</td>
</tr>
<tr>
<td>6</td>
<td>0.1883</td>
<td>0.7630</td>
<td>0.4795</td>
<td>0.1317</td>
<td>0.3657</td>
<td>0.0339</td>
<td>0.2059</td>
</tr>
<tr>
<td>7</td>
<td>0.4753</td>
<td>0.7815</td>
<td>0.4386</td>
<td>0.2850</td>
<td>0.4054</td>
<td>0.2059</td>
<td>0.3657</td>
</tr>
<tr>
<td>8</td>
<td>0.0694</td>
<td>0.3657</td>
<td>0.0209</td>
<td>0.0833</td>
<td>0.2059</td>
<td>0.3173</td>
<td>0.7055</td>
</tr>
<tr>
<td>9</td>
<td>0.4827</td>
<td>0.4795</td>
<td>0.2482</td>
<td>0.2482</td>
<td>0.5271</td>
<td>1.0000</td>
<td>0.1797</td>
</tr>
<tr>
<td>10</td>
<td>0.2876</td>
<td>0.3657</td>
<td>0.0522</td>
<td>0.1655</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.3173</td>
</tr>
</tbody>
</table>
5.2 Subjective Measures

Hypothesis 3 states subjects will be more satisfied with the solutions resulting from Feedback Iteration Mode or Iteration Mode than the solutions produced by Modification Mode or Manual Mode. Questions 7 and 10 in Table 2 were tested for subject satisfaction with the final solution provided by the systems. However, the p-values in the first column in Table 4 resulting from running an Omnibus Friedman show no statistically significant difference among the four different systems for the two questions. Moreover, we ran a pair-wise Friedman test, which showed no statistically significant differences for any pair of systems.

Looking at the histograms in Figures 5-8, we see some patterns that can help explain the results. The histogram for the manual mode in Figure 5 has two distinct peaks at 2 and 4. From the text feedback gathered at the end of each questionnaire, we hypothesize that there are two types of users. First, there are users who do not mind the cognitive load of the problems and were able to quickly develop strategies to effectively use the freedom provided by the Manual Mode to solve the problems effectively. The other group thought the cognitive load was too high and was not able to develop a strategy fast enough to solve the problems effectively without the help of an algorithm.

The histogram for Iteration Mode in Figure 7 also shows 2 peaks at 2 and 4. However, the relative heights are inverted. We hypothesize that this is due to the phenomenon we previously explained, where the group that was able to quickly develop effective strategies preferred Manual Mode and the others preferred having the algorithm do the computational part of the problem.

Hypothesis 4 states that subjects will report a higher workload for the Manual Mode and Modification Mode. Results from running a Friedman test on the first 6 questions of the post-scenario questionnaire in Table 2, show a statistically significant difference for question 1. The pairwise Friedman Test shows significance between Manual Mode and Feedback Iteration Mode in reported Mental Demand. This was expected since Manual Mode requires the subject to perform temporal and ergonomics
optimization, whereas Feedback Iteration Mode requires the subject to only perform ergonomics optimization. No statistically significance results support Hypothesis 5, which states that the subject will report higher trust for Iteration Mode and Feedback Iteration Mode.

5.3 Objective Measures

Hypothesis 2 states that schedules resulting from Iteration Mode and Feedback Iteration Mode will be objectively better than the schedules resulting from Manual Mode and Modification Mode. Before running statistical tests, we replaced makespan scores from invalid solutions with 1000 units of time, which was the maximum allowed time in the graphical user interface. Running a Repeated Measures Anova Test revealed no statistically significant difference across subjects or modes. However, looking at Table 18 and Figures 17-29, we see a trend that using Iteration Mode and Feedback Iteration Mode resulted in better makespans on average and similar ergonomic scores.
Chapter 6

Discussion

The aim of this work was to understand the effectiveness of different systems in 1) reducing the workload of the users, 2) producing objectively better solutions, 3) gaining the trust of the users, and 4) including user preferences. In this section, we discuss the results presented in the last section and consider future work that could build on these results.

6.1 Reducing Workload

The subjects reported less mental workload in Question 1 of the post-questionnaire for Iteration Mode and Feedback Iteration Mode. This was expected since Tercio was aiding in the scheduling and allocation of the tasks. However, the small difference in median score, 4.0 for Manual Mode and Modification Mode vs Iteration Mode and Feedback Iteration Mode, suggests the effect was slight. A plausible explanation is that most of the mental workload came from the temporal constraints, which were treated equally in all modes. The subjects reported in the open-ended part of the questionnaire that they found it difficult to properly visualize the temporal constraints. Recall that the only visual aid for temporal constraints was the coloring of tasks that were violating constraints. However, the constraints that were not satisfied were not highlighted. Furthermore, when multiple tasks violated constraints, the subjects were not always able to decipher the chain of temporal constraints. In
future work, better visualization of temporal constraints can lead to a reduction in mental workload.

6.2 Producing Objectively Better Solutions

Although the tests showed no statistically significant results supporting that the Iteration Mode and Feedback Iteration Mode produce objectively better schedules, we saw trends suggesting that these modes produce better solutions on average. For future work, including the ability to manually change the allocation instead of forcing users to run Tercio might result in stronger results. This is due to the fact that some subjects struggled to get the combination of allocation and sequence they desired since Tercio erased their sequence every time it ran.

6.3 Gaining the Trust of the Users

In [7], the authors state that an algorithm that provides an explanation of a solution is less likely to lose the trust of the user. We hypothesized that users would report higher trust in Feedback Iteration Mode. However, no statistically significant results support this hypothesis. While examining the open ended section, we notice that only three subjects read the feedback provided by the system. In future work, including feedback in forms other than text might result in more users taking advantage of the feedback.

6.4 Including User Preferences

The statistical tests showed no significant effect of the systems on the cumulative ergonomic scores. However, we think that including a live numerical score for the user preferences might not properly model real situations. User preferences that can be directly and easily quantified can most of the time be included in the problem formulation for the algorithm. In future work, giving the users information representing
the preferences but not providing them with numerical feedback might model the real scheduling environment more accurately.
Chapter 7

Conclusion

There is a need for scheduling interactive scheduling systems that allow users to include their preferences or domain specific knowledge. We developed a graphical user interface, which allowed the users to solve scheduling problems using different levels of automation and feedback. A 16-participant experiment was conducted to study the effectiveness of four different systems in producing solutions that optimize for makespan and for cumulative ergonomic score, a proxy for human preferences. Results showed statistically significant evidence (p=0.0027) that subjects reported higher workload when working with Manual Mode and Modification Mode rather than Iteration Mode and Feedback Iteration Mode. Finally, we proposed changes to our current system design and experimental design to guide future experimentation in interactive scheduling systems.
Appendix A

Tables
Table A.1: Build Car Agent Capabilities and Ergonomic Scores

<table>
<thead>
<tr>
<th></th>
<th>Human 1 Capability</th>
<th>Human 1 Ergonomic Score</th>
<th>Human 2 Capability</th>
<th>Human 2 Ergonomic Score</th>
<th>Robot Capability</th>
<th>Robot Ergonomic Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fetch Hood</td>
<td>50</td>
<td>80</td>
<td>45</td>
<td>80</td>
<td>44</td>
<td>0</td>
</tr>
<tr>
<td>Install Hood</td>
<td>70</td>
<td>35</td>
<td>80</td>
<td>40</td>
<td>90</td>
<td>0</td>
</tr>
<tr>
<td>Fetch Door 1</td>
<td>45</td>
<td>55</td>
<td>45</td>
<td>55</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>Fetch Door 2</td>
<td>45</td>
<td>55</td>
<td>45</td>
<td>55</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>Install Door 1</td>
<td>60</td>
<td>45</td>
<td>70</td>
<td>45</td>
<td>65</td>
<td>0</td>
</tr>
<tr>
<td>Install Door 2</td>
<td>60</td>
<td>45</td>
<td>70</td>
<td>45</td>
<td>65</td>
<td>0</td>
</tr>
<tr>
<td>Fetch Trunk Lid</td>
<td>60</td>
<td>75</td>
<td>65</td>
<td>75</td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td>Install Trunk Lid</td>
<td>90</td>
<td>50</td>
<td>100</td>
<td>50</td>
<td>110</td>
<td>0</td>
</tr>
<tr>
<td>Fetch Tire 1</td>
<td>40</td>
<td>35</td>
<td>45</td>
<td>35</td>
<td>45</td>
<td>0</td>
</tr>
<tr>
<td>Fetch Tire 2</td>
<td>40</td>
<td>25</td>
<td>45</td>
<td>25</td>
<td>45</td>
<td>0</td>
</tr>
<tr>
<td>Fetch Tire 3</td>
<td>40</td>
<td>35</td>
<td>45</td>
<td>35</td>
<td>45</td>
<td>0</td>
</tr>
<tr>
<td>Fetch Tire 4</td>
<td>40</td>
<td>25</td>
<td>45</td>
<td>25</td>
<td>45</td>
<td>0</td>
</tr>
<tr>
<td>Install Tire 1</td>
<td>50</td>
<td>55</td>
<td>55</td>
<td>55</td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td>Install Tire 2</td>
<td>50</td>
<td>70</td>
<td>55</td>
<td>70</td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td>Install Tire 3</td>
<td>50</td>
<td>55</td>
<td>55</td>
<td>55</td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td>Install Tire 4</td>
<td>50</td>
<td>70</td>
<td>55</td>
<td>70</td>
<td>70</td>
<td>0</td>
</tr>
</tbody>
</table>

Table A.2: Build Car Delays

<table>
<thead>
<tr>
<th></th>
<th>Releasing Task</th>
<th>Constraint Task</th>
<th>Delay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fetch Hood</td>
<td>Install Hood</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>Fetch Door 1</td>
<td>Install Door 1</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Fetch Door 2</td>
<td>Install Door 2</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Fetch Trunk Lid</td>
<td>Install Trunk Lid</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>Fetch Tire 1</td>
<td>Install Tire 1</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Fetch Tire 2</td>
<td>Install Tire 2</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Fetch Tire 3</td>
<td>Install Tire 3</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Fetch Tire 4</td>
<td>Install Tire 4</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Install Hood</td>
<td>Install Tire 1</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>Install Hood</td>
<td>Install Tire 2</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>Install Hood</td>
<td>Install Tire 3</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>Install Hood</td>
<td>Install Tire 4</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>Install Tire 1</td>
<td>Install Door 1</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Install Tire 2</td>
<td>Install Door 2</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Install Hood</td>
<td>Fetch Tire 3</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Install Hood</td>
<td>Fetch Tire 4</td>
<td>50</td>
<td></td>
</tr>
</tbody>
</table>
### Table A.3: Build Car Deadlines

<table>
<thead>
<tr>
<th>Releasing Task</th>
<th>Constraint Task</th>
<th>Deadline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Install Hood</td>
<td>Fetch Tire 4</td>
<td>250</td>
</tr>
<tr>
<td>Install Hood</td>
<td>Fetch Tire 3</td>
<td>250</td>
</tr>
</tbody>
</table>

### Table A.4: Build Desk Agent Capabilities and Ergonomic Scores

<table>
<thead>
<tr>
<th></th>
<th>Human 1</th>
<th>Human 2</th>
<th>Robot</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Capability</td>
<td>Ergonomic Score</td>
<td>Capability</td>
</tr>
<tr>
<td>Fetch Legs</td>
<td>80</td>
<td>80</td>
<td>90</td>
</tr>
<tr>
<td>Install Leg 1</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>Install Leg 2</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>Install Leg 3</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>Install Leg 4</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>Position Desk 1</td>
<td>90</td>
<td>90</td>
<td>100</td>
</tr>
<tr>
<td>Fetch Drawer 1</td>
<td>60</td>
<td>30</td>
<td>70</td>
</tr>
<tr>
<td>Fetch Drawer 2</td>
<td>60</td>
<td>30</td>
<td>70</td>
</tr>
<tr>
<td>Install Drawer 1</td>
<td>60</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Install Drawer 2</td>
<td>60</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Position Desk 2</td>
<td>100</td>
<td>80</td>
<td>100</td>
</tr>
<tr>
<td>Fetch Top</td>
<td>50</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>Install Top</td>
<td>90</td>
<td>90</td>
<td>80</td>
</tr>
</tbody>
</table>

### Table A.5: Build Desk Delays

<table>
<thead>
<tr>
<th>Releasing Task</th>
<th>Constraint Task</th>
<th>Delay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fetch Legs</td>
<td>Install Leg 1</td>
<td>60</td>
</tr>
<tr>
<td>Fetch Legs</td>
<td>Install Leg 2</td>
<td>60</td>
</tr>
<tr>
<td>hFetch Legs</td>
<td>Install Leg 3</td>
<td>60</td>
</tr>
<tr>
<td>Fetch Legs</td>
<td>Install Leg 4</td>
<td>60</td>
</tr>
<tr>
<td>Position Desk 1</td>
<td>Install Drawer 1</td>
<td>40</td>
</tr>
<tr>
<td>Position Desk 1</td>
<td>Install Drawer 2</td>
<td>40</td>
</tr>
<tr>
<td>Fetch Drawer 1</td>
<td>Install Drawer 1</td>
<td>40</td>
</tr>
<tr>
<td>Fetch Drawer 2</td>
<td>Install Drawer 2</td>
<td>40</td>
</tr>
<tr>
<td>Fetch Top</td>
<td>Install Top</td>
<td>50</td>
</tr>
<tr>
<td>Position Desk 2</td>
<td>Install Top</td>
<td>50</td>
</tr>
<tr>
<td>Install Drawer 1</td>
<td>Position Desk 2</td>
<td>70</td>
</tr>
<tr>
<td>Install Drawer 2</td>
<td>Position Desk 2</td>
<td>70</td>
</tr>
</tbody>
</table>
### Table A.6: Build Desk Deadlines

<table>
<thead>
<tr>
<th>Releasing Task</th>
<th>Constraint Task</th>
<th>Deadline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fetch Legs</td>
<td>Install Leg 1</td>
<td>300</td>
</tr>
<tr>
<td>Fetch Legs</td>
<td>Install Leg 2</td>
<td>300</td>
</tr>
</tbody>
</table>

### Table A.7: Build House Agent Capabilities and Ergonomic Scores

<table>
<thead>
<tr>
<th></th>
<th>Human 1</th>
<th>Human 2</th>
<th>Robot</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Capability</td>
<td>Ergonomic Score</td>
<td>Capability</td>
</tr>
<tr>
<td>Fetch Bricks</td>
<td>50</td>
<td>80</td>
<td>60</td>
</tr>
<tr>
<td>Fetch Cement</td>
<td>50</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>Fetch Sand</td>
<td>50</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>Fetch Gravel</td>
<td>60</td>
<td>90</td>
<td>70</td>
</tr>
<tr>
<td>Fetch Door</td>
<td>60</td>
<td>45</td>
<td>70</td>
</tr>
<tr>
<td>Fetch Window</td>
<td>40</td>
<td>45</td>
<td>50</td>
</tr>
<tr>
<td>Fetch Door</td>
<td>40</td>
<td>75</td>
<td>50</td>
</tr>
<tr>
<td>Make Concrete</td>
<td>80</td>
<td>50</td>
<td>90</td>
</tr>
<tr>
<td>Build Wall 1</td>
<td>100</td>
<td>35</td>
<td>110</td>
</tr>
<tr>
<td>Build Wall 2</td>
<td>100</td>
<td>25</td>
<td>110</td>
</tr>
<tr>
<td>Build Wall 3</td>
<td>100</td>
<td>35</td>
<td>110</td>
</tr>
<tr>
<td>Build Wall 4</td>
<td>100</td>
<td>25</td>
<td>110</td>
</tr>
<tr>
<td>Install Window</td>
<td>70</td>
<td>55</td>
<td>80</td>
</tr>
<tr>
<td>Install Door</td>
<td>70</td>
<td>60</td>
<td>80</td>
</tr>
</tbody>
</table>

### Table A.8: Build House Delays

<table>
<thead>
<tr>
<th>Releasing Task</th>
<th>Constraint Task</th>
<th>Delay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fetch Bricks</td>
<td>Make Concrete</td>
<td>60</td>
</tr>
<tr>
<td>Fetch Sand</td>
<td>Make Concrete</td>
<td>70</td>
</tr>
<tr>
<td>Fetch Gravel</td>
<td>Make Concrete</td>
<td>60</td>
</tr>
<tr>
<td>Fetch Cement</td>
<td>Make Concrete</td>
<td>70</td>
</tr>
<tr>
<td>Fetch Window</td>
<td>Install Window</td>
<td>0</td>
</tr>
<tr>
<td>Make Concrete</td>
<td>Build Wall 1</td>
<td>40</td>
</tr>
<tr>
<td>Make Concrete</td>
<td>Build Wall 2</td>
<td>40</td>
</tr>
<tr>
<td>Make Concrete</td>
<td>Build Wall 3</td>
<td>40</td>
</tr>
<tr>
<td>Make Concrete</td>
<td>Build Wall 4</td>
<td>40</td>
</tr>
<tr>
<td>Fetch Door</td>
<td>Install Door</td>
<td>0</td>
</tr>
<tr>
<td>Build Wall 1</td>
<td>Install Door</td>
<td>60</td>
</tr>
<tr>
<td>Build Wall 2</td>
<td>Install Window</td>
<td>60</td>
</tr>
</tbody>
</table>
### Table A.9: Build House Deadlines

<table>
<thead>
<tr>
<th>Releasing Task</th>
<th>Constraint Task</th>
<th>Deadline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Make Concrete</td>
<td>Build Wall 1</td>
<td>300</td>
</tr>
<tr>
<td>Make Concrete</td>
<td>Build Wall 2</td>
<td>300</td>
</tr>
</tbody>
</table>

### Table A.10: Build Airplane Agent Capabilities and Ergonomic Scores

<table>
<thead>
<tr>
<th></th>
<th>Human 1</th>
<th></th>
<th>Human 2</th>
<th></th>
<th>Robot</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Capability</td>
<td>Ergonomic Score</td>
<td>Capability</td>
<td>Ergonomic Score</td>
<td>Capability</td>
<td>Ergonomic Score</td>
</tr>
<tr>
<td>Fetch Wing 1</td>
<td>60</td>
<td>60</td>
<td>70</td>
<td>60</td>
<td>44</td>
<td>0</td>
</tr>
<tr>
<td>Fetch Wing 2</td>
<td>60</td>
<td>60</td>
<td>70</td>
<td>60</td>
<td>90</td>
<td>0</td>
</tr>
<tr>
<td>Fetch Propeller</td>
<td>40</td>
<td>70</td>
<td>50</td>
<td>70</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>Fetch Fuselage</td>
<td>80</td>
<td>50</td>
<td>90</td>
<td>50</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>Fetch Stabilizer</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td>Position Wing 1</td>
<td>60</td>
<td>80</td>
<td>60</td>
<td>80</td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td>Install Wing 1</td>
<td>80</td>
<td>50</td>
<td>80</td>
<td>50</td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td>Position Wing 2</td>
<td>60</td>
<td>80</td>
<td>70</td>
<td>80</td>
<td>110</td>
<td>0</td>
</tr>
<tr>
<td>Install Wing 2</td>
<td>80</td>
<td>50</td>
<td>80</td>
<td>50</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>Install Propeller</td>
<td>50</td>
<td>40</td>
<td>50</td>
<td>40</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>Position Fuselage</td>
<td>60</td>
<td>90</td>
<td>70</td>
<td>90</td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td>Install Fuselage</td>
<td>70</td>
<td>30</td>
<td>70</td>
<td>30</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>Install Stabilizer</td>
<td>50</td>
<td>30</td>
<td>60</td>
<td>30</td>
<td>70</td>
<td>0</td>
</tr>
</tbody>
</table>

### Table A.11: Build Airplane Delays

<table>
<thead>
<tr>
<th>Releasing Task</th>
<th>Constraint Task</th>
<th>Delay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fetch Wing 1</td>
<td>Position Wing 1</td>
<td>60</td>
</tr>
<tr>
<td>Fetch Wing 2</td>
<td>Position Wing 2</td>
<td>60</td>
</tr>
<tr>
<td>Fetch Propeller</td>
<td>Install Propeller</td>
<td>0</td>
</tr>
<tr>
<td>Fetch Fuselage</td>
<td>Position Fuselage</td>
<td>50</td>
</tr>
<tr>
<td>Fetch Stabilizer</td>
<td>Install Stabilizer</td>
<td>70</td>
</tr>
<tr>
<td>Position Wing 1</td>
<td>Install Wing 1</td>
<td>20</td>
</tr>
<tr>
<td>Position Wing 2</td>
<td>Install Wing 2</td>
<td>20</td>
</tr>
<tr>
<td>Position Fuselage</td>
<td>Install Fuselage</td>
<td>60</td>
</tr>
<tr>
<td>Install Fuselage</td>
<td>Position Wing 1</td>
<td>70</td>
</tr>
<tr>
<td>Install Fuselage</td>
<td>Position Wing 2</td>
<td>70</td>
</tr>
</tbody>
</table>
### Table A.12: Build Airplane Deadlines

<table>
<thead>
<tr>
<th>Task</th>
<th>Deadline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position Wing 1</td>
<td>180</td>
</tr>
<tr>
<td>Install Wing 1</td>
<td>180</td>
</tr>
<tr>
<td>Position Wing 2</td>
<td>180</td>
</tr>
<tr>
<td>Install Wing 2</td>
<td>180</td>
</tr>
</tbody>
</table>

### Table A.13: Median Scores for Post-Questionnaire

<table>
<thead>
<tr>
<th>Question</th>
<th>Manual</th>
<th>Modification</th>
<th>Iteration</th>
<th>Feedback</th>
<th>Iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.0000</td>
<td>4.0000</td>
<td>3.0000</td>
<td>3.0000</td>
<td>3.0000</td>
</tr>
<tr>
<td>2</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>3</td>
<td>4.0000</td>
<td>3.0000</td>
<td>3.5000</td>
<td>3.0000</td>
<td>3.0000</td>
</tr>
<tr>
<td>4</td>
<td>3.0000</td>
<td>3.0000</td>
<td>3.0000</td>
<td>3.0000</td>
<td>3.0000</td>
</tr>
<tr>
<td>5</td>
<td>3.5000</td>
<td>4.0000</td>
<td>3.0000</td>
<td>3.0000</td>
<td>3.0000</td>
</tr>
<tr>
<td>6</td>
<td>3.0000</td>
<td>3.0000</td>
<td>3.0000</td>
<td>2.0000</td>
<td>2.0000</td>
</tr>
<tr>
<td>7</td>
<td>2.0000</td>
<td>2.5000</td>
<td>4.0000</td>
<td>4.0000</td>
<td>4.0000</td>
</tr>
<tr>
<td>8</td>
<td>3.0000</td>
<td>3.5000</td>
<td>4.0000</td>
<td>4.0000</td>
<td>4.0000</td>
</tr>
<tr>
<td>9</td>
<td>5.0000</td>
<td>4.0000</td>
<td>4.0000</td>
<td>4.0000</td>
<td>4.0000</td>
</tr>
<tr>
<td>10</td>
<td>5.0000</td>
<td>4.0000</td>
<td>4.0000</td>
<td>3.0000</td>
<td>3.0000</td>
</tr>
</tbody>
</table>

### Table A.14: Median Objective Measures

<table>
<thead>
<tr>
<th>Median Scores</th>
<th>Manual</th>
<th>Modification</th>
<th>Iteration</th>
<th>Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ergonomics</td>
<td>410</td>
<td>457.5</td>
<td>415</td>
<td>445</td>
</tr>
<tr>
<td>Time</td>
<td>578</td>
<td>537</td>
<td>540.5</td>
<td>523.5</td>
</tr>
<tr>
<td>Total</td>
<td>1019</td>
<td>1012</td>
<td>962.5</td>
<td>975.5</td>
</tr>
</tbody>
</table>
Table A.15: P values resulting from running Wilcoxon Signed-Rank Test on Subjective Measures

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.3633</td>
<td>0.0181</td>
<td>0.0039</td>
<td>0.1191</td>
<td>0.1133</td>
<td>1.0000</td>
</tr>
<tr>
<td>2</td>
<td>0.2500</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.5000</td>
<td>0.2500</td>
<td>1.0000</td>
</tr>
<tr>
<td>3</td>
<td>0.0234</td>
<td>0.0996</td>
<td>0.0045</td>
<td>0.6777</td>
<td>0.5898</td>
<td>0.3047</td>
</tr>
<tr>
<td>4</td>
<td>0.5039</td>
<td>0.4767</td>
<td>0.1370</td>
<td>1.0000</td>
<td>0.4316</td>
<td>0.2578</td>
</tr>
<tr>
<td>5</td>
<td>0.8828</td>
<td>0.6719</td>
<td>0.0566</td>
<td>0.8271</td>
<td>0.0859</td>
<td>0.2764</td>
</tr>
<tr>
<td>6</td>
<td>0.7002</td>
<td>0.2891</td>
<td>0.0303</td>
<td>0.6807</td>
<td>0.0469</td>
<td>0.2832</td>
</tr>
<tr>
<td>7</td>
<td>1.0000</td>
<td>0.1863</td>
<td>0.1117</td>
<td>0.2686</td>
<td>0.1016</td>
<td>0.7051</td>
</tr>
<tr>
<td>8</td>
<td>0.3486</td>
<td>0.0352</td>
<td>0.0225</td>
<td>0.2168</td>
<td>0.2852</td>
<td>0.9844</td>
</tr>
<tr>
<td>9</td>
<td>0.4297</td>
<td>0.1997</td>
<td>0.5342</td>
<td>0.8633</td>
<td>0.5186</td>
<td>0.2500</td>
</tr>
<tr>
<td>10</td>
<td>0.1279</td>
<td>0.1074</td>
<td>0.0342</td>
<td>0.9121</td>
<td>0.8110</td>
<td>0.4844</td>
</tr>
</tbody>
</table>

Table A.16: P-Values resulting from running Repeated Measures Anova on Objective Measures

<table>
<thead>
<tr>
<th>Effect</th>
<th>Subjects P-value</th>
<th>Algorithm P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ergonomic Scores</td>
<td>0.2774</td>
<td>0.0298</td>
</tr>
<tr>
<td>Makespan P-Value</td>
<td>0.0573</td>
<td>0.6110</td>
</tr>
<tr>
<td>Total Score P-Value</td>
<td>0.1044</td>
<td>0.4682</td>
</tr>
</tbody>
</table>
Appendix B

Figures

Figure B-1: Answers to question 7 after using Manual mode. Median: 2

Figure B-2: Answers to question 7 after using Modification mode. Median: 2.5
Figure B-3: Answers to question 7 after using Iteration mode. Median: 4

Figure B-4: Answers to question 7 after using Feedback Iteration mode. Median: 4

Figure B-5: Answers to question 7 after using Manual mode. Median: 2
Figure B-6: Answers to question 7 after using Modification mode. Median: 2.5

Figure B-7: Answers to question 7 after using Iteration mode. Median: 4

Figure B-8: Answers to question 7 after using Feedback Iteration mode. Median: 4
Figure B-9: Total Score for Manual Mode

Figure B-10: Total Score for Modification Mode

Figure B-11: Total Score for Iteration Mode
Figure B-12: Total Score for Feedback Iteration Mode
Bibliography


