Cognitive Telepresence in Human-Robot Interactions

by

Vahagn Harutyunyan

A Thesis Presented to the

Masdar Institute of Science and Technology

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in

Computing and Information Science

Cognitive Telepresence in Human-Robot Interactions

By Vahagn Harutyunyan

A Thesis Presented to the Masdar Institute of Science and Technology in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Computing and Information Science

August 2012

AUTHOR'S DECLARATION

I understand that copyright in my thesis is transferred to Masdar Institute of Science and Technology.

Author ______________________________

RESEARCH SUPERVISORY COMMITTEE MEMBERS

Dr. Jacob Crandall, Chair, ______________________________
Masdar Institute of Science and Technology

Dr. Iyad Rahwan, ______________________________
Masdar Institute of Science and Technology

Dr. Zeyar Aung, ______________________________
Masdar Institute of Science and Technology

Abstract

Remotely operated semi-autonomous robots have great potential in many applications important to the economy and the environment in the Middle East. Applications include construction and maintenance of nuclear power plants and underwater oil wells, remote sensing for weather forecasting and pollution detection, and healthcare. Such tasks often require an operator without substantial technical expertise to remotely control a complex robot that is situated in an unknown and rapidly changing environment. In such case, the limited autonomy of the robot can cause failures that endanger the successful and efficient completion of the mission.

In this research, we argue that cognitive telepresence (CT), that is, the ability of the user to comprehend and control the robot's cognition, is central to the reliability and robustness of robot's autonomy, which in turn, significantly increases the chance of success. Correcting the robot's behavior without switching to lower autonomy levels (manual teleoperation) will allow better exploitation of the currently available automation algorithms.

To validate the applicability of our approach, we conducted experiments with human subjects. The subjects directed a simulated semi-autonomous robot in a scenario that models a real-world application. The task was to disarm a minefield without entering into any of several threat zones in the environment. The simulated robot had all the major capabilities required for accomplishing the task. However, the robot's autonomy was imperfect, as it sometimes misclassified mines and failed

to detect threat zones in the environment. Thus, the users had to be able to correct the robot's cognition to achieve mission success. This scenario, although hypothetical, illustrates the challenges that modern robotic applications must address in the real world.

The tests were conducted on five systems that differed from each other only by the levels of CT achieved by the users. The results of the experiment indicate that CT as a design feature of human-robot interaction systems can improve the overall performance of human-robot teams. It can also reduce the workload of the user and, without any changes to the robots' artificial intelligence, allow the completion of tasks that were unattainable with systems that provide lower levels of CT.

This research was supported by the Government of Abu Dhabi to help fulfill the vision of the late President Sheikh Zayed Bin Sultan Al Nayhan for sustainable development and empowerment of the UAE and humankind.

Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

1.1 Motivation

Robots are becoming common elements of various scientific, industrial, and medical systems, including deep-sea operations, space exploration, surgery, search and rescue, reconnaissance and other remote-sensing applications. Because of factors related to economy, the environment, and the society, robots will likely find serious application in the Gulf Cooperation Council (GCC) countries. The main industries of many GCC countries are developing around export of energy carriers such as oil and gas. Remotely operated vehicles (ROVs) are currently being applied in such tasks, and it is anticipated that they will be applied in off-shore projects as they mature [14, 21, 36].

Concerns related to the economy and the environment have affected the plans for power generation in GCC countries. The Abu Dhabi government has made a decision to construct four nuclear power plants by the year of 2020, which is expected to cover approximately one fourth of their power consumption [37]. Robots can be applied in a variaty of tasks related to the operation and maintenance of such objects. As an example, robots were applied during the Fukushima-Daiichi nuclear crisis.

Socially, autism affects a large number of people around the world and in GCC countries. The record suggests that the use of robots could be an effective tool in autism therapy [10, 20, 29, 30, 31]. While applying robotics in this domain, one has to deal with many challenges, including the design of interactive autonomy and easy-to-use interaction interfaces that would allow novice users to operate complex semi-autonomous robots.

Robots used for the above mentioned applications will differ in many ways. However, we expect that they all will have several common characteristics. Firstly, the robots will operate in dynamically changing environments the features of which will not be clear to the designers *a priori*. Secondly, the users might ask the robot to perform tasks that the robot was not designed or tested for, but which the robot is physically capable of performing. Lastly, the robots will need to have some autonomous capabilities since, given the complexity of their hardware and artificial intelligence (AI), they would otherwise require too much input from the operator in case of manual control.

Figure 1.1 portrays the dependencies between these three factors. If the characteristics of the environment and the tasks are bounded, sufficient AI can be designed and applied. Often, in such tasks, supervisory control is applied [34]. In that case, the robot is able to complete the assigned mission. The users simply supply high-level commands, and supervise the robots during task execution. In this context, the predictability of autonomous behaviors of robots becomes crucial. However, modern robotic autonomous control algorithms are very complex and consist of both reactive and planning elements that, given complex and dynamic

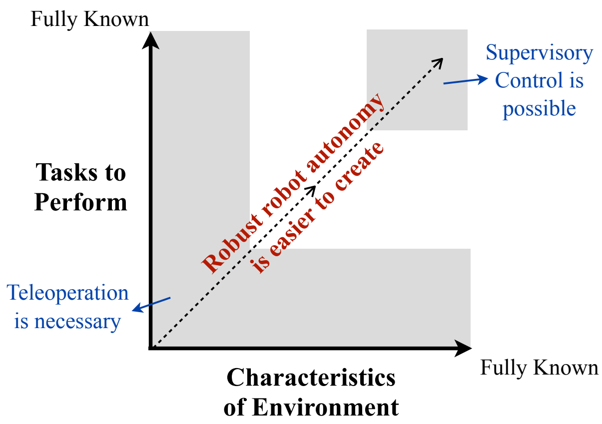


Figure 1.1: As designers of robotic systems have more knowledge of the task and environment, it becomes easier for them to create robust robot autonomy.

environments, make even the near-term behavior of the robot hard to predict. In addition, the operator who interacts with a semi-autonomous robot often lacks information about the internal (or cognitive) state of the robot, which includes its beliefs and intentions. This raises questions about the way the robot interprets various elements of the environment, and the tasks assigned by the operator. This, in turn, causes difficulties for the operator to identify causes of discrepancies between assigned and actually executed tasks of the robot. Moreover, even in case of identification of those causes, typical HRI interfaces do not provide substantial provisions for the operator to interact with the robot's cognition.

In this thesis, we aim at identifying the design features for HRI systems that affect the ability of users to understand and modify the cognitive state of the robot. For doing so, we first define the concept of Cognitive Telepresence (CT) and Correctability of Cognition (CorC) and illustrate methods for evaluating different systems in terms of those concepts.

1.2 Objective

The goal of this thesis is to investigate the possibility of transferring the information about a robot's cognitive processes to the operator in an easily interpretable form, and to investigate how HRI interfaces can be improved to increase the ability of the operator to modify those processes efficiently. That would allow the operator to identify possible inconsistencies between the instructions given to the robot and the actual behavior of the robot, and to make it easier for the operator to remove these undesirable inconsistencies. Human-robot teams that have such properties are expected to operate in a highly coordinated and cooperative manner. To address the above-mentioned issues, we define the concepts of Cognitive Telepresence (CT) and Correctability of Cognition (CorC) which are expected to reflect the user's ability to comprehend and modify the robot's cognition in an efficient manner. We also propose possible metrics that would allow system designers to identify the effects of various CT-related features in the relative performance of different HRI systems. Our goals are:

1. Identifying to what extent and in which condition does the ability of the user to comprehend and modify the cognition of the robot affect the performance of the overall HRI system

2. Investigating how having comprehension and/or control over one or another aspect of robot's cognition influences the user's performance and interaction effort.

To answer the above questions, we conducted a user study where subjects interacted with a simulated Minesweeper robot via various user interfaces.

1.3 Thesis Organization

Chapter 2 provides background information about human-robot interactions, their characteristics, and common metrics used for evaluating them. In Chapter 3, we define Cognitive Telepresence, its elements, and it's relationship to the well-known concepts in human-robot systems. Chapter 4 presents the proposed metrics for assessing cognitive telepresence in different systems. The case study we perfomed for validating the concept and the metrics is desribed in Chapter 5. Finally, in Chapters 6 and 7 respectively, we present the results of the user study and then draw conclusions.

CHAPTER 2

Literature Review

In this chapter, we provide background information for the thesis. Different concepts common in the field of robotics and human-robot interaction are discussed. We also present several metrics for assessing performance of robotic systems and their users while interacting with those systems.

2.1 Common Concepts in HRI

"Human-Robot Interaction (HRI) is a field of study dedicated to understanding, designing, and evaluating robotic systems for use by or with humans" [12]. HRI has many categories and subcategories, which differentiate between the distance and type of interactions between the robot and the human.

Teleoperation is the activity of directly controlling a distant robotic agent in continuous time. Supervisory control in turn is the process of controlling a robotic agent who has autonomous capabilities or is "automated" [34]. Autonomy of an agent is defined as the competence of sustaining the ability to reach the goal in case of a dynamically changing environment [23]. Sheridan [34] also defines ten levels of autonomy, which range from direct teleoperation to full autonomy. Alternatively, HRI systems are often characterized in terms of mixed-initiative strategy [13]. This scale differentiates the levels of interaction between a human and a robot, ranging from teleoperation to peer-to-peer collaboration. It has been argued that the level of autonomy should be adjustable [6], and the agent should be able to act fully autonomously in some situations, even without permission of the operator [16].

To assign a desirable level of autonomy to a robot, one should be aware of the situation in which the robot is operating including both the properties of the environment, and the status of the robot and its cognitive processes. There are several concepts that define the operator's knowledge about the current state of the remote process and her ability to modify the state from distance. Some of the commonly used terms are: telepresence, situation awareness, common ground, and fluency in interaction.

Telepresence is defined as "the case where a person is objectively present in a real environment that is physically separate from the person in space" [32]. In the ideal case, there should be full telepresence, i.e. the operator should have all the feelings which she would have had if she were actually interacting with the environment. However, Sheridan [34] argues that, in many cases, the use of teleoperation is motivated by the fact that the operator needs to be isolated from effects of some properties of the environment. These properties can include sudden peaks of resistant forces, vibrations, and high temperatures. Hence, the achievement of full telepresence is not necessarily a goal in the design of most HRI interfaces.

Situation awareness (SA) is defined to be "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" [8]. Basically, SA is a measure of information about the current status of the environment available to the operator, which is often provided in human-robot teams by video, audio, and haptic feedbacks. A discussion of SA in teleoperation is given by Riley et al. [28]. The authors claim that the efficiency of a teleoperation interface is highly correlated to the level of SA provided by it.

Telepresence and SA have been widely used in teleoperation and supervisory control. However, we believe that the concepts of common ground and fluency are more suitable for describing the interaction of an operator with the cognitive aspects of a robot. Common ground is the case when multiple agents share truthful mental models of each other [3]. In that case, they put the least collective effort while interacting with each other [19]. Kiesler claims that a human will interact with a robot in an efficient manner if they both have reasonable mental models of each other. This will in turn increase the predictability of robotic behaviors. Stubbs et al. [38] discuss the effects of increasing the level of autonomy on common ground in the context of a remote scientific exploration robot. They conclude that as the robot's autonomy increases, the problems related to common ground between the operator and the robot change from insufficiencies in context information and feedback to insufficient transparency. They claim that, for insuring a sufficient level of common ground, the robot should be able to adapt to the operators, their mental models, as well as to model the operators' high level tasks as they command the robot to perform lower level tasks.

Another concept for describing the quality of HRI is fluency. It is defined to be the case when two agents are operating in a well-coordinated manner and are habituated to each other and the mission [15]. Hoffman and Breazeal experimented with a behavior-based robot involved in a collocated interaction with humans, who were selected arbitrarily. The goal of the experiment was to investigate the effects of using a cognitive architecture implanted in the robot which provided the ability of early prediction of operators' moves. They show that the existence of such a cognitive module can dramatically decrease response times of the robot and hence speed up the completion of the task.

In this section, we summarized background information about the concepts present in the field from which we will build a more comprehensive concept. In general, these concepts are related to information flow in HRI systems. Telepresence is the measure of how much information the operator has about a remote robot's environment. Situation awareness defines the operator's knowledge about the current state of the environment and her ability to predict the near future. Common ground and fluency are more related to cognitive processes that define the behavior of the robot from the operator's point of view, and the habits of the operator from the robot's point of view.

2.2 Common Metrics in HRI

There are several metrics that describe the quality of an HRI system. All the concepts described in the previous section are measured with qualitative and/or quantitative metrics.

Several metrics for evaluating HRI interfaces in the framework of supervisory control are presented in [5], which defines neglect tolerance and interface efficiency for quantifying the autonomous abilities of a robot and the effectiveness of a HRI interface correspondingly. Neglect tolerance relates to the time period of acceptable performance which the robot can sustain while it is neglected by the operator. Neglect tolerance is different for different levels of autonomy. For example, in case of teleoperation (no autonomy), the effectiveness will dramatically decrease when the robot is neglected for even a short period of time, while in case of full autonomy, the robot should sustain a constant performance level regardless of the operator's activity. Interface efficiency is quantified by the time which the operator has to spend on accomplishing the full process of changing the state of the robot via that interface. This includes gaining sufficient SA, planning the command, and transmitting it to the robot.

These metrics can be used to evaluate an HRI scheme in general, without being concerned with detailed analysis of HRI elements, namely SA, telepresence, common ground, fluency, or CT.

A description of a technique called SAGAT for determining the SA of an aircraft pilot in a simulation is given in [8]. At some random point of the flight, the simulation is stopped and all the displays stop providing any information. The measure of SA is the number of questions about the current situation correctly answered by the pilot. This metric while having several disadvantages is still one of the most accepted in various fields.

Schloerb [32] describes several techniques for measuring telepresence. The author classifies telepresence to be either objective or subjective, and discusses methods for measurement of both. A person is objectively present only if she is able to accomplish the mission. Schloerb also distinguishes between different types of objective presence and argues that there are different degrees of telepresence in respect to each of the tasks. Degree of objective telepresence is roughly the probability of completing the task. As for subjective telepresence, the author defines it to be the case when a person has a feeling of being present. This is obviously harder to measure. The measurement in this case is done by asking the operator qualitative questions which allows comparing different interfaces.

Common ground is harder to measure quantitatively. We have not seen such a metric defined in the HRI literature. However, the concept itself is very relevant to the topic of the current thesis work since it considers the mutual understanding of each others thoughts between different agents (people and robots in our case).

Hoffman and Breazeal [15] used some primitive metrics for measuring fluency in HRI. They compared fluencies in different experiments in terms of time metrics, including overall task completion time, and mean sequence attempt time. This can hardly be considered a sufficient method for measuring fluency since the concept covers a much broader scope of properties of intelligent agent interactions.

After the discussion of common HRI metrics, we can conclude that there is a large space for research in the field; Namely, in the identification of elements of interactions that relate to people's understanding of the robots' current activities, their intentions, and the reasons for those intentions that is encompassed by CT. Hence, we attempt to develop metrics that would complement the use of CT in the pursuit of a better understanding of HRI, and the development of effective interfaces in the field of intelligent agents' interactions in general.

CHAPTER 3

Cognitive Telepresence

There exist two methods for human operators to overcome unreliable or insufficient robot autonomy. The first option is switching to manually controlling the robot until it is able to proceed independently. Second, the users can affect the robot's decision in some way without reducing the robot's level of autonomy.

In the first case, the user will need a sufficient understanding of the robot's physical state (e.g., orientation and position) and the conditions of the environment (e.g., the slope and close obstacles). These aspects are captured by SA. Also, the user will need to have a sufficient ability to simultaneously control each of the robot's functions.

In the second case, the user will need to understand the part(s) of robot's cognition that cause the error. For example, an object that was expected to be picked-up was classified as an obstacle, and hence, the robot avoided it. To do this, the user will need information about the robot's cognition, including its world model, perceived objectives of the mission, and the near and long term decisions. The user

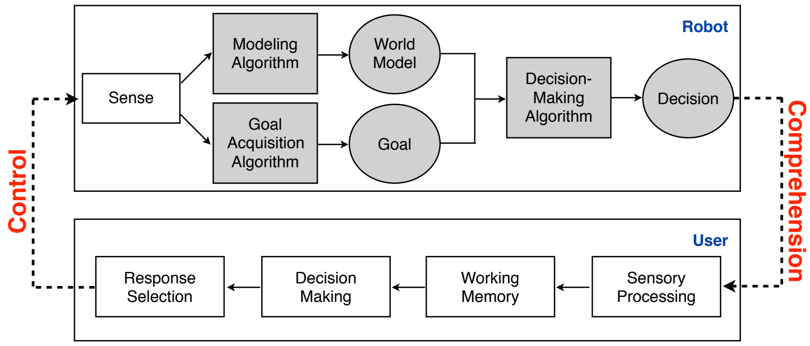


Figure 3.1: Cognitive telepresence refers to the user's *comprehension* and *control* of the robot's cognition (shaded shapes).

will also need to be provided with the ability to make modifications to the robot's cognition by the interface. These two features of the interaction scheme[1] affect the users' Cognitive Telepresence.

Definition: Cognitive telepresence (CT) is the user's **comprehension** of and **control** over the robot's cognition.

CT is expressed in Figure 3.1, which represents the cognitive aspects of human-robot interaction. The user is provided with information about the robot's cognitive processes (boxes) and states (circles). Based on that information, the user decides whether changes are needed in any of the states or processes and communicates them when necessary. In case of an identified discrepancy between the robot's cognition and the cognition the user believes the robot should have, the user can modify the particular cognitive aspect according to the amount of control she is provided with by the interaction scheme. We refer to the two components of CT as *comprehension of the robot's cognition* (CompCo) and *control over the robot's cognition* (ConCo). We describe them separately.

[1]An interaction scheme is the combination of the human-robot interface and the robot's artificial intelligence.

3.1 Comprehension of the Robot's Cognition (CompCo)

CompCo describes how much of the robot's cognition is perceived by the user. The cognition can be thought of as the state, and the processes that the robot performs to derive that state.

3.1.1 Elements of CompCo

As can be seen in Figure 3.1, we break down the robot's cognitive state into three elements: its model of the environment (*world model*), its *goal*, and the actions it plans to take (i.e., its *decision*). Meanwhile, its corresponding cognitive processes constitute the way the robot derives its world model and goals from its inputs (*modeling algorithm* and *goal acquisition algorithm*), and how it makes its decisions from its world model and goals (*decision-making algorithm*).

The robot's world model is derived by processing its inputs by its modeling algorithm. For example, in a navigation task, the robot's world model might consist of a map of the world, which might include labels of items in the world and an obstacle map. The user's CompCo of the world model refers to the degree to which the user knows what the robot's world model is. Similarly, the user's CompCo of the robot's modeling algorithm refers to the user's understanding of how the robot's world model is derived from its inputs. For example, the user might understand that the robot is not classifying a particular item correctly because the item does not meet a precoded color threshold in the camera frame.

The robot's goal is formulated by processing its inputs using its goal acquisition algorithm. Therefore, CompCo with respect to the robot's goal refers to how well the user knows what the robot is trying to accomplish. For example, in a navigation task, CompCo of the robot's goal can be conveyed by displaying the robot's intentions in case of identifying another robot. Possible goals include starting to follow the peer or keeping distance from it. Similarly, the user's CompCo of the robot's goal acquisition algorithm refers to the user's understanding of how the robot's goal was formed from its inputs.

Finally, the robot's decision is determined by the decision-making algorithm given the robot's world model and goal. CompCo with respect to the robot's decision refers to how well the user knows the actions the robot plans to take to accomplish its goal. In a navigation task, CompCo of the robot's goal can potentially be established by marking a path on a map display indicating to the user the route the robot intends to follow to its destination. Similarly, the user's CompCo of the robot's decision-making algorithm refers to the user's ability to understand how and why the robot made the decision.

An example of a mechanism designed to (partially) establish CompCo of the decision-making algorithm is provided in a study conducted by Thomaz and Breazeal [39]. In this study, the users were teaching a simulated robot how to bake a cake. The robot learned, using Q-learning, from the user's input, which the user was giving while observing the robot's activity. In one of the scenarios, the robot paused and glanced toward desirable actions when it was confused between them. This helped the user to understand that the robot placed similar Q-values on actions, and that it did not know which action to take. In this way the users were able to understand the way the robot's decision making worked.

3.1.2 Contributing Factors

To provide CompCo for each of the cognitive elements, a human-robot interface must satisfy two necessary conditions: *observability* and *clarity*. In the control theory literature [17, 18], observability refers to the presence of information provided, in our case, to the user. For the user to possibly have complete CompCo, the HRI interface must display as much information about the robot's cognition as is necessary for her to be able to completely model the robot's thoughts.

However, observability alone can not guarantee CompCo. The user must also properly process the information available through the human-robot interface. Thus, the second interface requirement (clarity) refers to the ease at which the user can interpret the information that the robot communicates about its cognition. Recalling the system created by Thomaz and Breazeal [39], a human-robot interface could have displayed the robot's Q-values associated with each action on the screen. However, a pause and glance from one action to the next provided a natural gesture that users easily understood as confusion.

3.2 Control of the Robot's Cognition (ConCo)

Control of the robot's cognition (ConCo) refers to how fast and which cognitive states and processes of the robot the user can modify. As with CompCo, we discuss the elements and contributing factors of ConCo.

3.2.1 Elements of ConCo

As with CompCo, ConCo is defined with respect to the robot's world model, goal, decision, modeling algorithm, goal acquisition algorithm, and decision-making algorithm. ConCo refers to the user's ability of efficiently modifying each element as she finds necessary. The amount of ConCo depends on how many of the cognitive aspects of the robot the user can change, and how fast she can do that.

Robotic systems described in the literature often provide users with high ConCo with respect to the robot's goals and decisions. For example, when a robot plans to select a path the user thinks is unsuitable, many robotic systems allow the user to assign different waypoints or modify the existing waypoints to change the robot's intended path. Nevertheless, such interactions can require substantial effort from users in the form of manual planning and implementation. We argue that providing the user with the ability of altering the robot's world model (such as communicating the unseen hazard to the robot) can result in simpler and more effective interactions that can better utilize the robot's limited autonomy. We have observed that such interactions and capabilities are less common in existing robotic systems. Moreover, when possible, allowing the user to make simple adjustments to the robot's modeling, goal acquisition, or decision-making algorithms can provide long-term improvements in the robot's autonomy.

We focus on these concepts by using an illustrative example in Chapter 5.

3.2.2 Contributing Factors

ConCo (for each cognitive element) is affected by two things: *controllability* and *expression*. Controllability, a term commonly used in control theory literature [25], refers to the power that the user has over each of the robot's cognitive elements. To have full ConCo of a given cognitive element, it must be possible for the user to efficiently change that robot's cognition (with respect to that element) as and when she finds necessary.

Nonetheless, as with observability, controllability alone does not guarantee ConCo. The second necessary factor is expression [26], which is similar in its meaning to the gulf of execution discussed by Norman [24]. Expression describes the user's ability to determine and communicate the input to the robot that has the desired effect on the robot's cognition. In the ideal case, such input will be natural such that the user can quickly communicate the desired modification to the robot.

3.3 Relationship of CT with Well-Known Concepts

CT is related to a number of well-known design principles commonly used in the literature on human-robot systems and introduced in Chapter 2. To better quantify

CT, we discuss its relationship with many of these principles, including situation awareness (SA), telepresence, common ground, levels of automation, adjustable autonomy, and neglect tolerance.

3.3.1 Situation Awareness

One of the widely used concepts in the field of human-robot interaction is situation awareness (SA). The term *situation awareness* was originally coined by the US military, particularly in regard to air combat. It has gained different meanings over time. Recall the definition by Endsley [8]. For single-robot systems, Drury et al. defined the term *HRI awareness* as “the understanding the human has of the location, activities, status, and surroundings of the robot” [7].

We think that these efforts have mainly focused on understanding the robot’s physical environment, its position in the environment, and the tasks the robot is carrying out, with some effort to begin to model the robot’s status, such as the condition of the robot’s battery [7]. CompCo is related, but is more specific. It refers particularly to the user’s awareness of the robot’s cognitive state and processes inasmuch as they implement some form of robot autonomy. Figure 3.2 illustrates the difference between SA and CompCo. SA refers to the user’s model of the world, while CompCo refers (in part) to the user’s model of the robot’s world model. Thus, CompCo could potentially be re-phrased as *situation awareness of the robot’s cognition*. However, even if we accept this definition, CT also refers to the additional attribute of control over the robot’s cognition, which SA certainly does not. In spite of the differences in the definitions, the main reasons for considering SA and CT are the same. The user must have both CT and SA to effectively interact with the robot so that it can complete the assigned mission. Together, the user’s SA and CT define the user’s ability to correct errors in the robot’s autonomy. Consequently, the ability of the user to correct the robot’s cognition, called

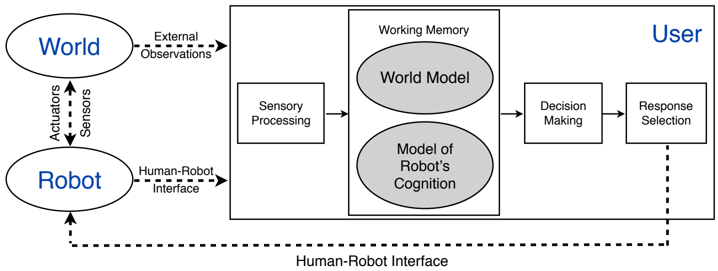


Figure 3.2: A human-robot interface should be designed so that a user can build an adequate *model of the world* (situation awareness) as well as a model of the robot's cognition (comprehension of cognition).

correctability of cognition (or CorC), is a function of SA and CT.

$$CorC = f(SA, CT). \tag{3.1}$$

We note an important relationship between SA and CT. Having high control over the robot's cognition with low SA means that users are able to modify the robot's cognition, but have less ability to modify it towards correction. Therefore, high CT in conjunction with low SA is not very useful. Similarly, high SA together with with low CT means that the user will likely have to switch to manual control to overcome errors in the robot's autonomy.

3.3.2 Telepresence

The notion of CT is also related in some respects to *telepresence*, defined as "the case where a person is objectively present in a real environment that is physically separate from the person in space" [32]. Similar to telepresence, CT refers to the feeling of sharing the same brain (rather than physical space), which involves being able to read and change thoughts. Sheridan argued that it is not always desirable for a user to have full telepresence in human-robot systems [34]. Likewise, too much CT can negatively impact the performance of a human-robot system.

3.3.3 Common Ground

The concept of *common ground* is also related to CT. Common ground, related in many ways to CompCo, refers to multiple entities sharing truthful mental models of each other [3]. That allows the entities to communicate effectively with each other with minimal effort [19].

Stubbs et al. [38] argue that, for providing sufficient common ground, a robot should be able to adapt to the user's mental model. This proposed solution is similar to ConCo, except that ConCo refers to the user's ability to change the robot's cognition rather than the robot's ability to adapt to the user's mental model.

3.3.4 Levels of Automation

Another related and widely referenced concept in HRI systems are the *levels of automation* [35]. Level of automation defines the interactions between the human and the machine while collectively performing a task. The highest level assumes fully autonomous operation by the robot, whilst the lowest, is the case when the human takes full control.

We perform comparisons between CT and levels of automation. First, in a sense, CT is exactly opposite to the concept of levels of automation. Levels of automation refer to the use of the machine to help the user complete a task, whereas CT addresses the problem of using humans facilitating the robot complete the task. As discussed by Bainbridge, robot autonomy simply alters the nature of human-robot interactions; it does not eliminate the need for them [1].

Second, CT can be an important aspect of any level of automation where manual control is not available. Regardless of the role being played by the robot, the user may want/need to comprehend and control the robot's cognition. However, we argue that, as the systems rely more and more on the robot's autonomy (i.e., higher levels of automation), CT becomes more crucial.

3.3.5 Adjustable Autonomy

In the literature a widely used approach for working around the failures in robot's autonomy is making the autonomy adjustable [33, 2, 22, 27]. Adjustable autonomy refers to the activity of switching (either at the choice of the user or the robot) to a different level of automation depending on the current situation. For example, in a navigation task, the robot might initially navigate autonomously. When the robot's autonomous navigation algorithm fails to find a way to navigate through a particular area of the world, the user could manually control the robot through the difficult area before returning the responsibility of navigation back to the robot.

Nevertheless, as discussed in the introduction, manual control of a complex robot in an unknown and complex environment can require extreme time and effort from the user, and it may not be even possible. A consideration of CT offers an alternative. Instead of switching to manual control (or some other low level of autonomy), the system can be designed so that the user can modify the robot's cognition, such as helping the robot to identify passable or impassable regions in the environment. Such approach could potentially offer a more effective solution that requires less effort from the user and is longer term.

3.3.6 Neglect Tolerance: Neglect and Interaction Times

The neglect tolerance is the frequency and duration of human-robot interactions that are required for the robot to maintain acceptable performance for a given interaction scheme [5]. In its simplest form, neglect tolerance is derived from the neglect time (the average amount of time the robot can be ignored by the operator before its performance drops below some threshold) and the interaction time (the average duration of a single human-robot interaction) [26, 11]. More comprehensive models of neglect tolerance use random processes to emphasize the trade-off between neglect times, interaction times, and average performance [4]. In any case,

Cognitive Element	**Impact on Short-term Performance**	**Impact on Future Neglect Times**	**Required Interaction Times**
World Model	Medium	Medium	Medium
Goal	Medium	Medium	Medium
Decision	High	Low	Low
Modeling Algorithm	Low	High	High
Goal Acquisition Algorithm	Low	High	High
Decision Algorithm	Low	High	High

Table 3.1: Anticipated consequences for implementing high CT of each cognitive element.

human-robot systems exploiting more effective interaction schemes typically have longer neglect times, shorter interaction times, and higher average performance than those employing less effective interaction schemes.

CT is an important contributor to neglect tolerance. Appropriate levels of CT will increase the neglect tolerance of an interaction scheme, while too little or too much CT will likely lower it. However, not all elements of CT are expected to have the same affect on neglect tolerance. Table 3.1 lists anticipated consequences for implementing high CT for each cognitive element. The table captures the idea that understanding and controlling the robot's cognitive processes (algorithms) will require more time but have longer term impact (neglect time) than understanding and controlling the robot's cognitive states.

Designers of human-robot systems must find the best trade-off between implementing high CT for each cognitive element. The aspects of the interface that assist the user in understanding and controlling the robot's cognition can have critical long-term benefits. Nevertheless, the users may spend too much time seeking to understand and control different cognitive elements, which consumes the resources needed to attend to the other tasks the user must perform (such as maintaining SA).

CHAPTER 4

Measuring Cognitive Telepresence

In this chapter we define explicit, partial, and implicit measures of CT. CT can be assessed explicitly via carefully controlled user experiments. However, because aspects of CT are mental processes, we cannot fully observe it. Since CT correlates to other concepts (e.g., Eq. 3.1) in a complected manner, it is difficult to measure. Therefore, we argue that implicit measures that are simple to obtain are, in many cases, more desirable than compete measures. Designers of human-robot systems should carefully choose from these potential metrics for measuring the CT of their systems in an effective manner.

4.1 Explicit Metrics

CT can be explicitly assessed in terms of CompCo and ConCo for each cognitive element. We discuss each in turn.

4.1.1 CompCo

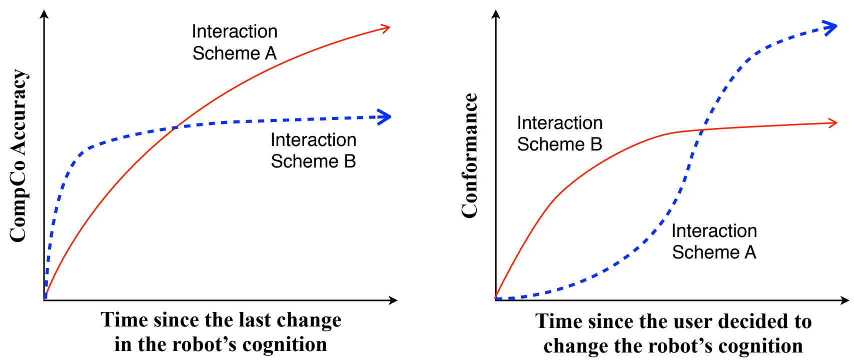
The CompCo of an interaction scheme can be defined as a random process that specifies the extent to which the user comprehends the robot's cognition (measured as a percentage of understanding) given the amount of time since the last change in the robot's cognition. The mean of this random process, which we define as the *CompCo accuracy curve*, can then be used to evaluate the interaction scheme's CompCo. A separate random process can be used to measure CompCo with respect to each cognitive element.

Figure 4.1a shows hypothetical CompCo accuracy curves for an arbitrary cognitive element for two interaction schemes. In scheme A, the user slowly gains CompCo as she interacts with the robot until she fully comprehends the robot's cognition. On the other hand, the user more quickly gains (perhaps adequate) levels of CompCo with scheme B, but then struggles to obtain higher levels of CompCo with additional time and effort.

Figure 4.1a illustrates the two important aspects of CompCo: time and accuracy. In general, effective interaction schemes provide the user with high CompCo more quickly than less effective interaction schemes. However, Figure 4.1a illustrates the potential trade-offs designers of human-robot systems face. Depending on system characteristics and needs, either interaction scheme A or B may be more desirable. Based on the CompCo accuracy curve $g(t)$ of a particular cognitive element, the CompCo with respect to that cognitive element is defined as:

$$CompCo = \frac{1}{\tau}\int_0^{\tau} g(t)\, dt, \tag{4.1}$$

where τ is the window of time over which the user should respond to the changes in the robot's cognition. Note that this method for measuring CompCo places weight on both the speed and accuracy of the user's comprehension of the robot's



(a) (b)

Figure 4.1: Hypothetical plots of the (a) CompCo accuracy and (b) ConCo conformance curves for two interaction schemes.

cognition.

We note that, as in other metrics associated with human cognition (such as situation awareness), measuring CompCo is not easy due to its interdependence on other aspects of the system. Nonetheless, CompCo accuracy could be measured by carefully controlled studies similar to those used to measure neglect tolerance [5].

4.1.2 ConCo

As is the case with CompCo, there are two relevant attributes that a good metric of ConCo should consider: (1) the amount of time it takes the user to modify the robot's cognition in the intended way, and (2) the ability of the user to control all the elements of the robot's cognition. As with CompCo, these attributes are captured in a random process that defines the conformance of the robot's cognition to the user's commands over time (indexed from the time the user decided to change the robot's cognition). We call the mean of this process the *ConCo conformance curve*.

Figure 4.1b shows hypothetical ConCo conformance curves for two interaction schemes. The figure illustrates the potential trade-off between long-term conformance and speed of control that designers of human-robot systems must balance.

ConCo for a given cognitive element can be computed by taking the integral of the ConCo conformance curve $h(t)$. Formally,

$$ConCo = \frac{1}{\tau}\int_0^{\tau} h(t)\, dt, \tag{4.2}$$

where τ is the window of time over which the user should respond to changes in the robot's cognition. As with CompCo, ConCo is not easily measured (but can be estimated via carefully constructed experiments) due to the difficulty of estimating the ConCo conformance curve $h(t)$.

4.2 Partial Metrics

Considering the difficulty of estimating CompCo and ConCo via the CompCo accuracy and ConCo conformance curves, respectively, simpler methods for measuring those elements of CT are necessary. While not fully indicative, these metrics can provide relative measures that would allow system designers to compare and contrast the CT of different interaction schemes.

4.2.1 CompCo

The CompCo of each cognitive element can be partially measured via two different metrics. First, CompCo can be partially measured by the observability of the cognitive element (Section 3.1.2). This metric simply considers the percentage of information about the cognitive element of the robot that the user could possibly identify via the human-robot interface. For example, if the robot's world model consists of the identified objects and their properties (such as type, color, size, etc.) recognized by the robot's modeling algorithm, then the observability of the world model is defined by the percentage of those properties that the user is informed about by the system. In a sense, the observability of a cognitive element represents the maximum possible CompCo (upper bound) of an interaction scheme for that particular cognitive element.

Second, as in the case of SA, the relative CompCo level for a cognitive element can be estimated by asking the user questions about the robot's cognitive states and processes during operation. Such questions can be administered by freezing the system and either blanking [8] or not blanking the screen [9].

4.2.2 ConCo

The upper bound on the ConCo for each cognitive element of an interaction scheme is the controllability of cognition (Section 3.2.2). Thus, controllability, expressed as a percentage of items in the robot's cognition that can possibly be modified, is an applicable partial metric of ConCo. Recalling the example of the robot's world model from the previous section, the controllability of the world model in that system would be defined by the percentage of the properties of objects in the world model that the user can modify via the interface.

Other metrics are necessary to determine how easily a user is able to communicate a command to the robot that will change the cognition in the desired way. This can be inferred by analyzing the user's behavioral patterns (such as observing mouse clicks) surrounding the changes to the robot's cognition made by the user.

4.2.3 Overall CT

Finally, we believe a useful metric for measuring overall CT is the correctability of cognition (CorC; see Eq. 3.1) provided by the human-robot interface. Given that CorC is a function of both CT and SA, it does not solely measure CT. However, when users have similar SA in each interaction scheme, CorC is a useful measure of CT, especially for relative comparisons between interaction schemes.

As with other metrics, CorC can be difficult to quantify. However a metric for assessing the relative CorC between systems can be applied. It involves considering the average amount of correction (difference in correctness) to the robot's cognition made by the user per unit of interaction time. The correctness of cognition is defined as the ratio of the cognitive elements that conform to the realities of the mission to the total number of those elements. For example the correctness of the robot's world model can be defined as the ratio of the correctly classified objects n_t and the total number of objects that are relevant to the mission N_t. In this case, the correctness of the robot's world model during a specific moment of time t can be defined by

$$Correctness(t) = \frac{n_t}{N_t} \tag{4.3}$$

The CorC with respect to the world model is defined as the average difference in the correctness made by the user per second of interaction. Formally,

$$CorC = \sum_{k=1}^{K} \frac{dn_k}{N_k t_k} \tag{4.4}$$

where K is the total number of interactions the user had during the mission, dn_k is the difference in the number of correctly classified objects before and after the interaction k, N_k is the total number of objects relevant to the mission in the environment during the interaction k, and t_k is the duration of interaction k.

We note that determining N_k (and perhaps even n_k) is somewhat subjective and is situation specific. However, a careful evaluation of each situation should allow for effective relative evaluations of CorC among interaction schemes.

4.3 Implicit Metrics

The ultimate goal of any human-robot system is high performance. Therefore, an essential metric of the effectiveness of a system with respect to CT is the performance of the system itself. We note that, in many cases, performance will be highly correlated with CorC, inasmuch as the N items used to compute CorC that are important to the mission success.

As referenced in Section 3.3.6, another implicit metric of CT is neglect tolerance, including neglect times and interaction times. Therefore, the impact system characteristics have on neglect times and interaction times indicate the way system characteristics affect CT.

CHAPTER 5

The Case Study

We conducted a user study to show how CT can be used as a design principle for human–robot systems. In this chapter, we describe the user study. We report the results of this study in Chapter 6.

5.1 Problem Description

In this study, users used a simulated robot to detect and disable land mines. The objectives were disarming all the mines in the shortest possible time and without entering designated threat zones. The environment consisted of two types of objects (mines and stones) and threat zones. Each day a new set of mines and stones was introduced, and additional threat zones were added, while keeping the old ones.

In the beginning of each day, the robot scanned the area for mines and stones. Afterwards, it classified the objects using the images taken while scanning. The classification was done considering two features of the image: color and regularity of the structure. The robot had default threshold values for each of the features and classified the objects accordingly. After classifying the scanned objects, the robot calculated the shortest path through all the objects that were classified as mines.

The robot was also able to detect some fraction of the threat zones, and was equipped with an algorithm that allowed it to avoid entering threats it knew about. This algorithm involved automatically setting an instant waypoint to guide the robot around the threat.

The robot did not always correctly classify the mines, stones, and threat zones. In case of the default classification threshold, the robot misclassified the mines as stones and vice versa around 20–30% of time. Also, it was able to detect only 10–15% of the threat zones. It also incorrectly assigned a threat to a safe area once in three days. In this situation, the subjects had to interact with the robot to keep it from failure.

This scenario shares several characteristics with tasks common to human-robot systems, including detection and classification of objects, navigation, obstacle avoidance, and repeated interaction with an environment that could potentially change over time. Real-world scenarios with such characteristics include maintenance of a nuclear power plant that has experienced a crisis, search and rescue, and remote sensing activities. The users interacted with the robot using a graphical user interface (GUI). A snapshot of the GUI is given in Figure 5.1. The GUI had the following features:

1. *Sensor display* – As is common in most human-robot interfaces, the system displayed the robot's sensor readings, namely, the robot's video stream was shown on the left-hand side of the GUI.

2. *Situation awareness display* – One feature of such human-robot interfaces that has become nearly universal is the "situation awareness" display. This

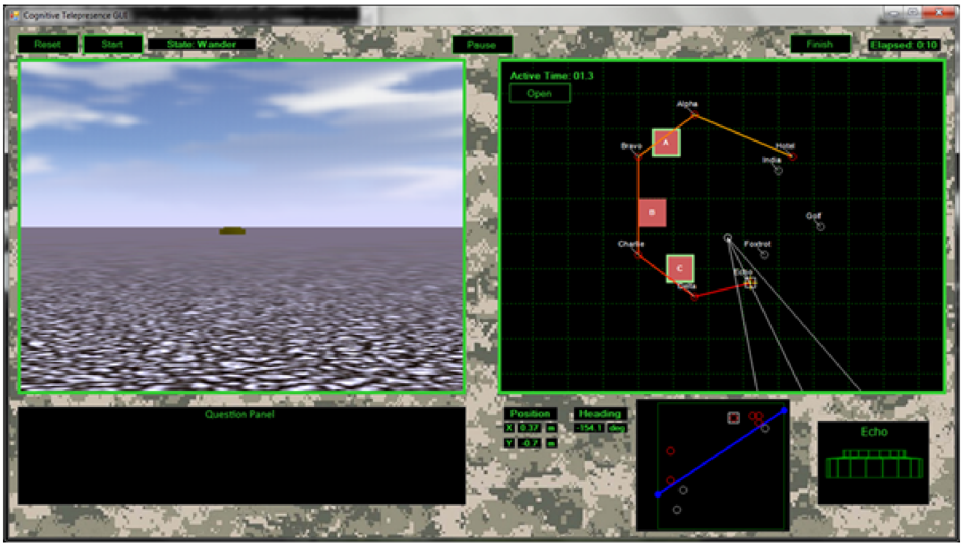


Figure 5.1: GUI used in the experiment

display is for providing the user with information about relative placement and orientation of the robot with respect to the surrounding objects. In our interface, the situation awareness display (right-hand side of Figure 5.1) showed the robot's current position in relation to the objects (circles) and hazards (squares) in the area. Additionally, the user was able to click on each object in the situation awareness display to view an image of that object that was taken by the robot while scanning the environment. This allowed the user to determine which objects were mines and which were stones with 100% certainty.

3. *Projected path* – The situation awareness display also showed a path indicating the current sequence of objects the robot planned to visit. Such displays have become common, and provide the user with comprehension over the robot's decision.

The above mentioned elements are important features of the system, however they do not provide the necessary amount of comprehension of and control over the robot's cognition. By using CT as design principle one is able to identify the additional information and capabilities that the user needs to be provided with by the interface to be able to effectively supervise a limited autonomy. In this case study, we consider improving the users' CT with respect to the robot's world model, modeling algorithm, and its short term decision as means for improving the system's performance. This focus on CT illustrates several potential improvements to the interface:

1. *Display the robot's object labels* – This feature, which was integrated into the situation awareness display in our interface, helps the user to comprehend the robot's world model (CompCo). All detected objects were shown in the situation awareness display. Objects that the robot believed to be mines were displayed in red, while objects the robot believed to be stones were displayed in gray.

2. *Indicate whether the robot is aware of threats* – This feature, which was also integrated into the situation awareness display in our system, also facilitated the user's comprehension of the robot's world model (CompCo). Threats the robot was aware of were highlighted by green square rings, while threats the robot was not aware of were not highlighted.

3. *Allow the user to change the robot's object labels* – This feature allowed the user to change the class of an object (either stone or mine) in the robot's world model. By right-clicking on an object and selecting the option "switch type", the user communicated the new class to the robot. The robot then automatically updated its world model and changed its path accordingly. This functionality provided the user with control over part of the robot's world model (ConCo).

4. *Allow the user to communicate the location of unseens threats to the robot*

– Since a robot may not be able to detect all threats that it should avoid, the capability to communicate those threats to the robot could be important if the robot is to rely on its autonomy to act. In our interface, the user simply right-clicked on the threat to inform about it the robot. Like the previously mentioned functionality, this provided the user with additional control over the robot's world model (ConCo).

5. *Allow the user to discard a threat area that the robot has detected* – This functionality is useful when the robot missclassifies a safe area as a threat zone. The user right-clicked inside the area that was designated to be a threat by the robot on the SA display and selected the option to discard the threat. This was another capability for the user to correct the robot's world model (ConCo).

6. *Allow the user to view and change the robot's classification threshold* – The robot classified the objects based on the color and structure of the objects using a linear threshold. In systems where this capability was active, the threshold as well as the location of each object in the feature space (color and regularity) was displayed below the situation awareness display. The user clicked on the endpoints of the classification line and modified its slope and displacement to the desired values. The robot then automatically re-classified the objects based on the new threshold, and updated its planned path accordingly. This display provided the user with some comprehension of and control over the robot's modeling algorithm.

The above-mentioned enhancements may or may not be considered obvious by skilled system designers. However, we argue that they are often absent in human-robot systems. By using CT as a design basis, system designers will be able to improve the systems so they can overcome problems related to the limitations in

Feature	*System 0*	*System 1*	*System 2*	*System 3*	*System 4*
CompCo *Planned Path*	○	●	●	●	●
CompCo *Object Labels*	○	●	●	●	●
CompCo *Known Threats*	○	○	○	●	●
ConCo *Object Labels*	○	○	●	●	●
ConCo *Notify Threats*	○	○	●	●	●
ConCo *Discard Threats*	○	○	○	●	●
CT *Classific. Algorithm*	○	○	○	○	●

Table 5.1: Attributes of five systems evaluated in the user study

robot autonomy.

5.2 Experimental Protocol

To demonstrate the effect of the differences in CT on the performance characteristics of the system, we conducted a user study using the human-robot system described in the previous section. In this study, we compared and contrasted five different variations of the system. The differences in the systems are outlined in Table 5.1.

Each of the systems had several basic features. The autonomous capabilities of the robot were the same through all the systems. However, only systems 2–4 allowed modifications to the robot's cognition. In all systems, the users were able to modify the robot's short term goal, i.e. set a particular object (regardless of the class assigned by the robot) as the current target. Also, the users were able to make the robot drive to a particular location by assigning instant waypoints. In systems 0 and 1, these two controls were the only ones available whenever the robot was selecting a stone as a target or was driving towards an undetected threat. The difference between those two systems was that system 0 did not provide the user with information about the results of classification of the objects by the robot. In other words, the user had to guess whether the robot considered a particular object as a stone or a mine. While providing the users with the additional ability to see the robot's interpretations of objects, system 1 provided the user with the same ConCo as system 0. The rest of the systems were enhanced by features that increased the observability of and controllability over the robot's cognition in multiple ways. The features were distributed through systems 2–4 according to the level of advancement that we anticipated. To explain, each new version of the system had one or more extra feature(s) that was/were absent in its predecessors, and each of those features was less essential but was intended to reduce the long-term workload more than the ones present in the predecessor systems.

System 2 provided the users with the ability to change the robot's model of the world. The users were able to both switch the classifications of objects in the robot's world model and inform the robot about undetected threats. However, the system did not show whether the robot knew about a particular threat or not. This system provided the users with some ConCo over the robot's world model, but it still did not provide full CompCo over it.

System 3 additionally provided the users with the ability to see which threats the robot knew about. Thus, it potentially allowed the users to have full information about the robot's world model. In this case, the users were also able to discard the threats in the safe zones that were identified to be dangerous by the robot. In other words, the users had full control over the robot's world model.

System 4 also allowed users to understand and modify the robot's modeling algorithm. Here, the users could change the threshold values that determined the values of color and regularity that indicated the classes of objects.

Each subject was tested on two systems on two different scenarios. A scenario consisted of three days. The scenarios differed by the placement of the objects and threats, but not by difficulty. In the beginning of each day, five to six mines and four to five stones were placed in the environment. In day 1, three threats were placed in the environment. The threats persisted from day 1 through day 3, but additional two threats were placed in day 2. The robot always knew about one of the threats unless the user had notified it about additional ones. However, in day 3, the robot always detected a safe zone around one of the mines as a threat zone, which caused additional difficulty for the user in identifying the robot's behavior. Thus, the difficulty levels were increasing from day 1 to day 3.

For system 4, the classification threshold, once set by the user, persisted throughout the scenario. The robot also remembered the locations of the threats from day 1 through day 3.

To assess the relative levels of SA and CompCo of the users, the users were asked questions that were designed to check their understanding of the situation in the environment and robot's cognition. The questions appeared under the camera image every 23 seconds after the last answer (or 23 seconds after the start) and caused the system to pause. To answer a question, the user had to click on a button labeled "Show the question" which allowed her/him to view that question. After that, the time and the number of clicks it took the user to answer the question was logged. Note that all the displays were available for the users when they were answering the questions.

SA was tested by asking questions about the objects (mines and stones) in the environment. For example, "Is object Alpha a rock or a mine?", where Alpha was the designation of an object on the SA display. The possible answers to this type of questions were: (1) Rock, (2) Mine, and (3) Unsure. For CompCo, there were two types of questions, both about the robot's world model, related to the classification of objects and the detected threats. First, “What does the robot think object Alpha is?”, with the same answers as the SA-related questions. Second, “Does the robot know about threat A?”, with possible answers (1) Yes, (2) No, and (3) Unsure. The variant “Unsure” was considered as a wrong answer, however the users were told that it is better to answer with “Unsure” rather than incorrectly. During each day, nine fixed questions (three of each type), specific to that day, were asked. We measured SA as the percentage of correct answers to SA-related questions. CompCo was measured as the average percentage of correct answers to both types of CompCo questions. For both SA and CompCo-related questions, we also measured the amount of time it took the participants to answer the questions as well as the number of clicks they made before supplying an answer.

Twenty Masdar Institute students between the ages of 23–32 with mean age of 26.6 participated in the study. Each subject were assigned a randomly selected pair of systems to operate. Table A.1 shows the resulting assignment of systems to the subjects.

For each participant, the following procedure was followed:

1. The participant completed a demographic survey.

2. The participant was trained on the first system until (s)he felt comfortable using the system.

3. The participant carried out the first test scenario using the first system, which lasted approximately 15–20 minutes.

4. The participant was trained on the second system until (s)he felt comfortable using this new system.

5. The participant carried out the second test scenario using the second system, which also lasted about 15–20 minutes.

CHAPTER 6

Results

In this chapter, we present the results of the measurements we did during the experiment. The goal was to find out whether having different levels of CT affects the performance of the robotic system and interaction times of the users. We focus on the relationships between the measures of CT, time-to-completion, the ratio of interaction time to total time, SA, and CorC.

6.1 Methodology

To evaluate the impact of the various system characteristics on performance and interaction times, we performed a series of statistical tests. First, for each metric, we compared the results of the systems by considering their means over all three days (7–8 data points for each system[1]). Since each subject was tested on two

[1] In the case of systems 1 and 4, due to a bug in the code, the data for day 3 was not logged for one of the runs. We discarded the results of all three days for those runs. Hence, the number of collected data points for systems 1 and 4 was 7, instead of 8.

systems, one could argue that the samples for these two systems were not independent. However, we treated them as independent samples for the statistical tests. We conducted one-way ANOVA tests on the mean values for all five systems to find whether there were statistically significant differences between them. For tests that showed statistically significant difference ($\alpha = .05$), we also performed pairwise t-tests to identify the pairs of systems that had statistically significant differences in the mean values of the particular metric. We also performed two-way ANOVA tests for metrics from both the systems and the days (5 systems and 3 days). This was done to identify whether there were significant effects that different days had on the values for systems. In this case as well, the significance was evaluated given the same assumptions as in the case with one-way ANOVA tests.

Finally, for better understanding the relationship between CorC and the other metrics, we analyzed the Pearson product-moment correlation coefficients between them. In this case, we did not differentiate between systems or days, but treated all points similarly. We considered the correlations with the same assumptions as with the ANOVA tests as described previously.

6.2 Performance

As previously mentioned, the objective that the subjects had to accomplish was to disarm as many mines as possible without entering any threat zone in minimal time. Thus, we will report the difference in the systems with respect to the number of mines missed, threats entered, and time to complete the mission.

The first performance metric we consider is the number of missed mines. Figure 6.1 shows that all the misses occurred in day 3. In this day, the robot thought that there was a threat zone around one of the mines. For systems that do not allow the operator to correct the robot's world model, the user was not able to force

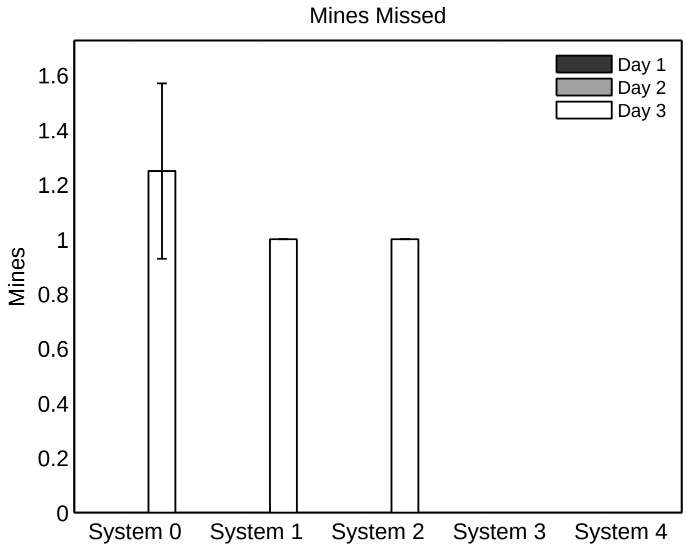


Figure 6.1: The number of mines missed

the robot to disarm that particular mine. Hence, the subjects had no choice but to finish the run without disarming it. However, since there was only one remaining mine, the users were able to disarm all the mines that were possible to disarm. Furthermore, notice that in systems 3 and 4 all the mines were disarmed. Recall that systems 3 and 4 in contrast to the others provided the users with the ability to discard threats and hence correct that aspect of the robot's cognition. In systems 1 and 2, the number of missed mines was exactly 1 in all the runs. However, in system 0, there was some deviation. Obviously, there was at least one mine that was not disarmed due to the same reason as in the case of systems 1 and 2. Hence, the deviation is a result of some subjects who failed to disarm two or more mines. This seems to be a consequence of either having less SA, or too much workload.

Our second performance metric relates to the entering of threat areas, which the users were told to avoid. We measure this in terms of the number of threats entered and the time spent in threats. Figure 6.2 shows a trend which indicates that systems with higher indices resulted in lower numbers of entered threats (F(4,33)

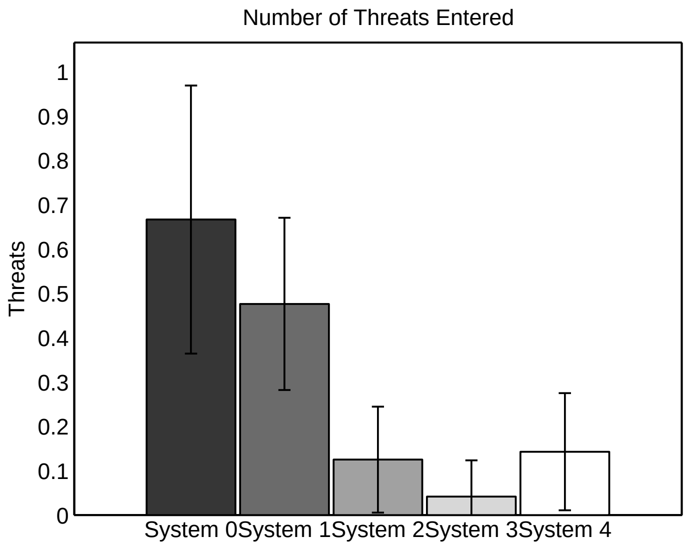


Figure 6.2: The number of threats entered

= 8.24, p < .001). Pairwise t-test showed that there was a statistically significant difference between system 0 and systems 2 (t(7) = 2.87, p = .023), 3 (t(7) = 4.25, p = .003), and 4 (t(6) = 3.03, p = .023). Also, there was a statistically significant difference between system 1 and systems 2 (t(7) = 2.87, p = .023), 3 (t(6) = 4.50, p = .004), and 4 (t(5) = 2.73, p = .040). In other words, when users were provided with the ability to notify the robot about the threat, the number of threats entered dropped substantially.

Figure 6.3 depicts the difference between the times in threats while using each system. An ANOVA shows a statistically significant difference between the systems (F(4,33) = 7.77, p < .001). Pairwise t-test shows significant difference between system 0 and systems 2 (t(7) = 3.86, p = .006), 3 (t(7) = 5.34, p = .001), and 4 (t(6) = 5.63 p = .001).

Figure 6.4 shows the time-to-completion for each system. An ANOVA shows a statistically significant difference between the systems (F(4,33) = 7.69, p < .001). Pairwise t-tests showed that system 4 was significantly different from systems 0

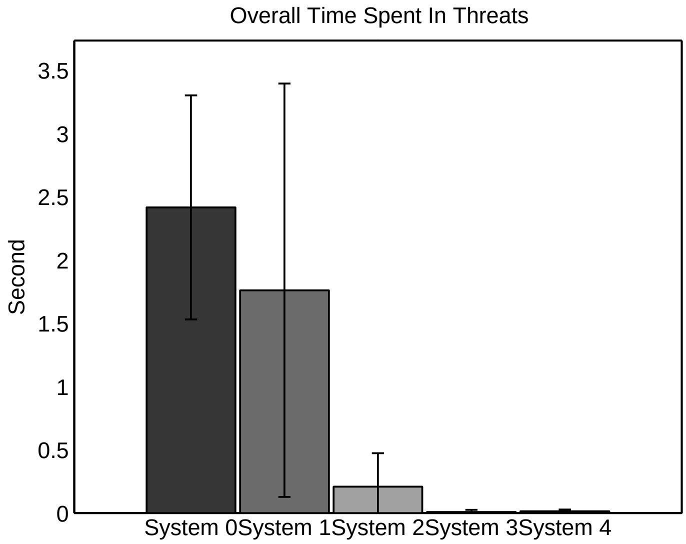


Figure 6.3: Overall time spent inside threats

(t(6) = 4.84, p = .003), 1 (t(5) = 3.86, p = .011), and 2 (t(6) = 2.46, p = .049). Additionally, system 3 significantly differed from systems 0 (t(7) = 6.33, p $<$ .001) and 1 (t(6) = 2.81, p = .030).

These results indicate a significant difference between the users' experiences with the various systems. Notice that the autonomy levels and the AI capabilities are the same across all the systems. The only things that were different were the amount of information that the users had and the number of cognitive elements of the robot that the users were able to modify.

6.3 Comprehension of Cognition (CompCo)

In this section, we present the results of the measurements that reflected users' CompCo. In our study, we measured CompCo as the percentage of correct answers to CompCo-related questions presented in the trials, and the time and effort spent on answering them for different systems. First, we consider the percentages

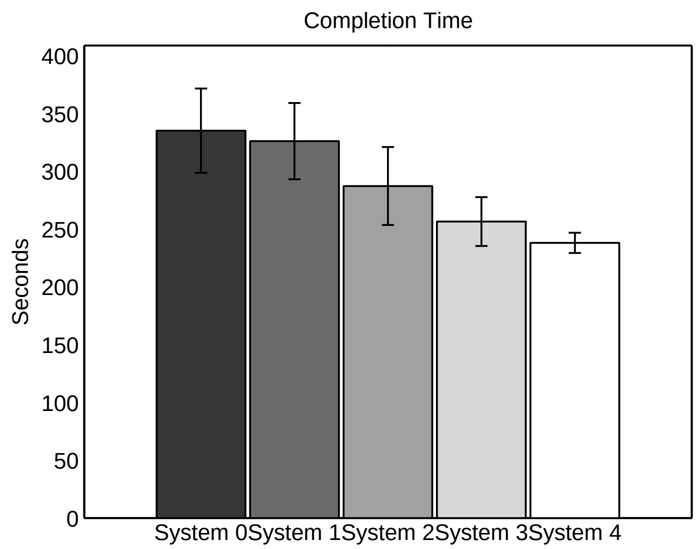


Figure 6.4: Completion times

of correct answers (Figure 6.5). An ANOVA shows statistically significant difference between the systems ($F(4,33) = 129.23$, $p < .001$). Pairwise t-tests show the following results. System 4 was significantly different from systems 0 ($t(6) = 4.84$, $p = .002$), 1 ($t(6) = 2.80$, $p = .030$), and 2 ($t(6) = 2.45$ $p = .049$). Also, system 3 was significantly different from systems 0 ($t(7) = 6.33$, $p < .001$) and 1 ($t(6) = 2.80$, $p = .030$).

From these results we can infer that, in our systems, the features of providing more or less information about the robot's cognition significantly impacted the users' actual ability to understand the robots' cognition. In other words, the differences in clarity that the interfaces of the systems provided the users with had less contribution on the user's CompCo than had the differences in observability (Section 3.1.2). This is the basis of our further comparisons and correlation analysis between the systems' CompCo, performance, and workload.

Now we consider the amount of time and effort (number of clicks) the users had spent on answering CompCo questions in different systems (Figure 6.6). An

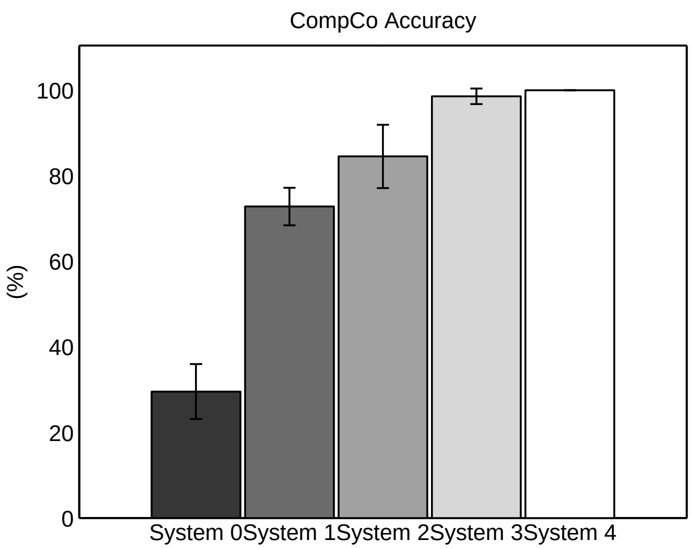


Figure 6.5: CompCo

ANOVA shows a significant difference in the amount of time for answering questions between the systems (F(4,33) = 2.78, p = .043). Pairwise comparisons showed that there was a significant difference between system 4 and systems 0 (t(6) = 2.86, p = .028), 1 (t(5) = 3.16, p = .025), 2 (t(6) = 2.79, p = .031), and 3 (t(6) = 3.07, p = .022). In summary, these results show that users were able to answer CompCo-related questions significantly faster with system 4 than with the rest of the systems.

The effort for answering CompCo-related questions measured in clicks is illustrated in Figure 6.7. An ANOVA does not show any statistically significant difference between the systems (F(4,33) = 0.34, p = .848).

6.4 Interaction Time

Interaction time is another commonly used metric in HRI. To estimate the differences in workload while using different systems, we measured the percentage of overall time the users spent on interacting with the robot, total number of interac-

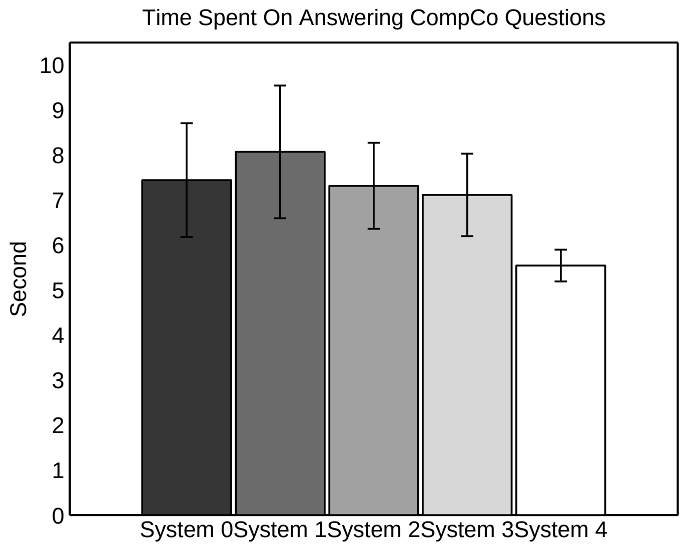


Figure 6.6: CompCo question answering times

tions per run, and the total number of clicks made during interactions.

The interaction percentage values are shown on Figure 6.8. An ANOVA test does not show any significant difference ($F(4,33) = 1.71$, $p = .172$). However, when we observe the interaction times per systems and days (Figure 6.9), we can see that while using system 4, the subjects had significantly less workload on days 2 and 3 than day 1. ANOVA test shows statistically significant difference between three days ($F(2,20) = 13.57$, $p < .001$). Pairwise t-tests performed on the values for all three days of system 4 showed that there was a significant difference between the values of interaction percentage between day 1 and days 2 ($t(7) = 3.76$, $p = .007$) and 3 ($t(6) = 4.58$, $p = .003$).

Recall that, in system 4, the users were provided with the ability to modify the robot's classification algorithm. This increased the workload in day 1. However, after finding good threshold for color and regularity for the classification algorithm and modifying the classification line, the users had much less to do in days 2 and 3 in comparison to day 1, as well as with the other systems. This illustrates our

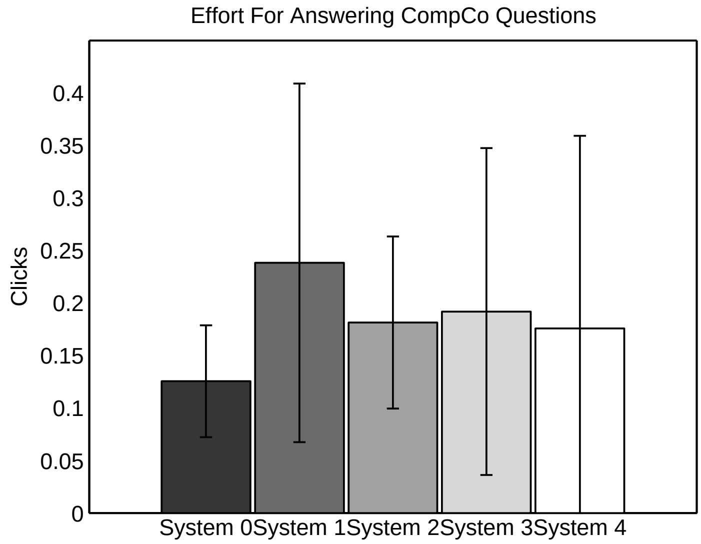


Figure 6.7: CompCo question answering efforts

hypothesis that the ability to change the robot's cognitive process can increase the short term workload while decreasing the long term one. We anticipate that if the number of days were larger for each system, than the total interaction percentage would have been significantly less in system 4 in comparison to the other systems.

6.5 Situation Awareness

In our experiment, we considered measuring the relative levels of SA. We measured the percentage of correctly answered SA-related questions as well as the time and effort that the subjects spent on answering them.

The data shows that all the subjects correctly answered all the questions besides one subject, who, while using system 0, once answered that he was unsure. Thus, with regard to this metric, there was no statistical difference in SA among the systems. Figure 6.10 shows the amount of time it took participants to answer SA-related questions in different systems. An ANOVA test shows no significant differ-

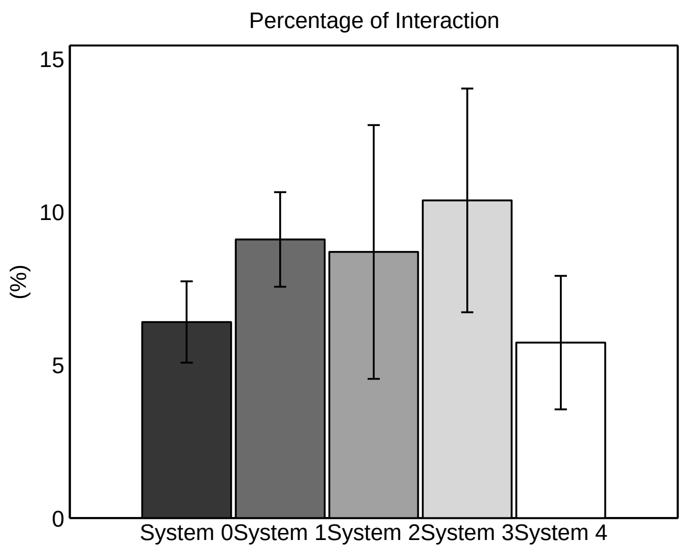


Figure 6.8: Interaction percentage

ence among the values for different systems ($F(4,33) = 1.51$, $p = .220$). The effort for answering SA questions did not differ from system to system either ($F(4,33) = 1.07$, $p = .388$). Neither did the values change significantly from day to day in any of the systems.

In conclusion, the levels of SA of all the subjects seemed to be the same. Therefore, we can argue that the difference in the performance, expressed by the number of missed mines, the number of entered threats, the total time spent inside threats, and interaction time, are primarily caused by differences in CT.

6.6 Correctability of Cognition

Correctability of the robot's cognition (CorC) is defined as the amount of correction of cognition per second of interaction by the user. CorC is an indirect metric for CT. We already discussed the results related to both CT and SA, which are the hypothetical components of CorC (section 4.2.3). We measured the average amount

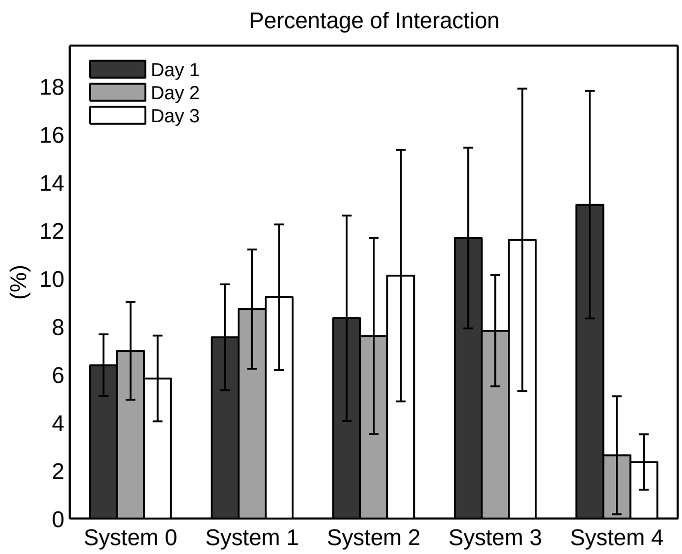


Figure 6.9: Interaction percentage by days

of correction of the robot's world model per second of interaction. We expected this metric to depict both the comprehension and, more importantly, the control of robot's cognition the subjects had while interacting with different systems.

As seen on Figure 6.11, the CorC in systems 0 and 1 was always 0. This is because, in those systems, the user could not modify the classification of objects or the number of threats known by the robot. All the control the users had over the robot was on a the short term decision. Namely, they were able to either set an instant waypoint or set an object as the current target. The values for the rest of the systems are a little harder to explain. Systems 2, 3, and 4 are significantly different from both system 0 ($t(7) = -6.90$, $p < .001$; $t(7) = -7.76$, $p < .001$; $t(6) = -6.73$, $p < .001$ correspondingly) and 1 ($t(6) = -6.46$, $p < .001$; $t(6) = -7.05$, $p < .001$; $t(5) = -5.80$, $p < .001$ correspondingly). However, there is no any other pairwise difference between systems.

For finding the effect of CorC on performance and workload we performed correlation analysis without discriminating by systems or days. This can be im-

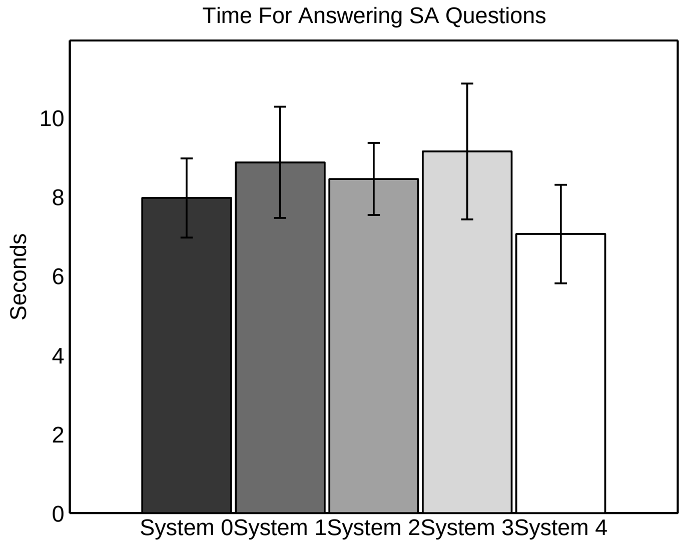


Figure 6.10: SA question answering time

portant, since it allows us to compare the values of CorC of all the runs at the same time without having to consider the averages by systems or days. We found that CorC was correlated with the percentage of interaction (r(118) = -0.28, $p < .01$). The total number of interactions and total interaction effort were also negatively correlated with CorC (r(118) = -0.47, $p < .01$ and r(118) = -0.43, $p < .01$ correspondingly).

As for performance, we found that CorC was negatively correlated with the number of missed mines (r(118) = -0.32, $p < .01$), the number of threats entered (r(118) = -0.26, $p < .01$), time spent in threats (r(118) = -0.28, $p < .01$), and the time-to-completion (r(118) = -0.46, $p < .01$).

We also tested the correlation between CorC and CompCo and found significant positive correlation (r(118) = 0.51, $p < .01$).

The above-mentioned correlations suggest that CorC may be a good measure of CT, since it correlates with both interaction time and system performance.

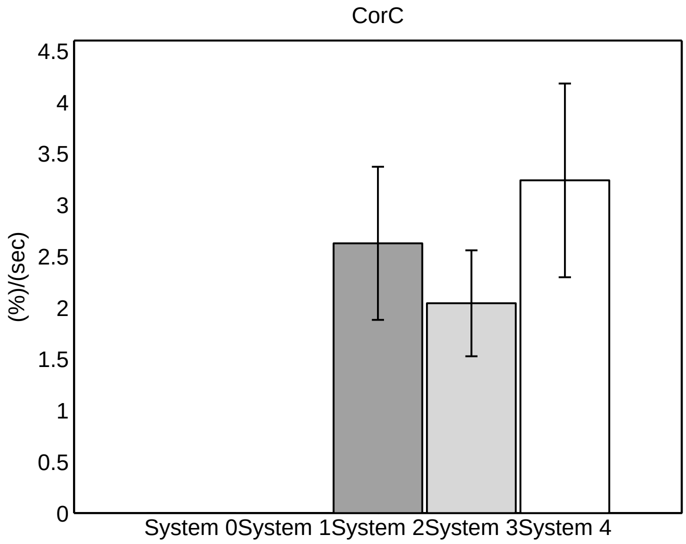


Figure 6.11: CorC

CHAPTER 7

Conclusion

As robotic systems mature, they are likely to come into high demand in GCC countries. Applications include robots for the oil and gas industries, power plant maintenance, and health care. Each of these applications is likely to share a set of common characteristics. First, the environment in which the robot operates in each robotic system will be somewhat noisy, dynamic, and unknown. Second, the tasks that users will want their robots to perform may not be fully known to system designers in advance. Third, the complexity of the robot will be such that some robot autonomy will be necessary to perform the desired tasks at a sufficient level of proficiency. Unfortunately, given the uncertainty in both tasks and environments, creating robust robot autonomy for such applications is extremely difficult. As a result, system designers must prepare for situations in which the robot's autonomy fails in unexpected ways.

In this research, we argued that cognitive telepresence (CT), or the ability of the user to comprehend and control the robot's cognition, is a critical design principle for building systems that can operate despite these failings. We identified six elements of a robot's cognition, three states and three processes, which the user should monitor and correct (if needed). The states are the robot's world model, its acquired goal, and its decision. The processes are the robot's modeling algorithm, its goal acquisition algorithm, and its decision-making algorithm. The robot observes the environment and builds a model of the world according to the modeling algorithm. Then, it acquires the task from the user and identifies its goals according to its goal acquisition algorithm. Finally, the robot makes the decisions based on the model of the world and the acquired goals, according to the decision-making algorithm. There can be failures in any of the above mentioned elements due to complex or unknown environments, assignment of tasks to the robot that were not considered in the robot's design, as well as failures in the hardware and software elements of the robot.

In case of having sufficient information about the cognitive elements of the robot, the user can identify the reasons for its undesired behavior, and even predict a failure in advance. The operator can also diagnose the problems in the cognitive states if she has proper understanding of the algorithms that result in those states. Also, for correcting the robot's cognition in real time, the user will need means of controlling the cognitive elements provided by the user interface. The amount of information that the user has about the robot's cognition is referred to as comprehension of the robot's cognition (CompCo), and the amount of control the user has over that cognition is referred to as control of the robot's cognition (ConCo).

To evaluate the effects of having different levels of CT on system performance and interaction time, we conducted a user study where the subjects interacted with a simulated semi-autonomous robot in a mine patrol task. The task was to disarm a minefield without entering into any threat area. The scenarios were designed in such a way that participants had to modify the robot's behavior to keep it from failure.

Five different systems were tested, all identical except the information about and control over the robot's cognition available for the users. All the systems provided the users with low level control over the short-term decisions of the robot. However, some of the systems allowed the user to understand and modify the robot's world model and decisions. One of the systems allowed the users to modify the robot's modeling algorithm as well.

The systems were tested in terms of system performance, interaction time, CompCo, situation awareness, and correctability of cognition.

As a measure of CompCo, we considered the percentage of correct answers to the CompCo-related questions given by the users during the experiment. The results show that the CompCo of the users who were interacting with the systems that provided information on a larger number of cognitive elements of the robot was substantially higher. Moreover, the systems that provided high CompCo showed better results in terms of the number of missed mines, the number of entered threats, and the time spent in threats by the robot.

The interaction times had no significant statistical difference throughout the systems. However, we observed that in the system where the users were able to affect the modeling algorithm, they had to put much less effort after correcting the modeling algorithm in comparison with the rest of the systems.

The measurement of the users' relative levels situation awareness was done by considering the percentage of correct answers to SA-related questions, as well as the time and the number of clicks it took them to answer those questions. The results imply that there was no significant difference in the levels of SA of the users throughout the systems. Based on this, we argue that the improvement in the performance was not due to the differences in the levels of SA of the users. Rather, it was due to the differences in users' comprehension and control over the robot's cognition.

For measuring overall CT in the systems indirectly, we considered estimating the values of Correctability of Cognition (CorC) provided by the systems. CorC is a function of SA and CT, and is defined to be the amount of correction to the robot's cognition the users make per second of interaction with the robot. Correctness of cognition is defined for each cognitive state and is the percentage of robot's cognition that corresponds to the the current situation. For example, in our case, we considered the correctness of the robot's world model. It was reflected in the percentage of correctly classified mines and known threats by the robot. While interacting with the robot, the users, depending on the particular system, had different amounts of control over this cognitive element. A statistical test shows that there is a strong correlation between the level of CorC and performance, interaction times, total number of interactions, and the total number of clicks the users made while interacting with the robot. Based on the observed absence of significant difference in the levels of SA across the systems, we argue that, in this case, CorC illustrates the differences in CT of the users across the systems. Hence, CT itself correlates to the above mentioned measures of performance and interaction time.

In conclusion, the performance of human-robot systems can be increased by providing the users with more information about and control over the robot's cognitive elements. In this case, the users are able to correct the robot's cognition instead of switching to manual control, and hence modify the robot's behavior in a more efficient manner.

APPENDIX A

The Schedule Of The Case Study

Subject	*Scenario 1*	*Scenario 2*
1	System 1	System 2
2	System 3	System 0
3	System 1	System 3
4	System 1	System 0
5	System 3	System 4
6	System 4	System 2
7	System 2	System 1
8	System 2	System 0
9	System 3	System 1
10	System 0	System 3
11	System 1	System 4
12	System 4	System 3
13	System 0	System 2
14	System 4	System 1
15	System 2	System 3
16	System 2	System 4
17	System 3	System 2
18	System 4	System 0
19	System 0	System 4
20	System 0	System 1

Table A.1: The distribution of the systems over the subjects

APPENDIX B

Abbreviations

HRI Human-Robot Interaction

CT Cognitive Telepresence

SA Situation Awareness

CompCo Comprehension of Cognition

ConCo Control of Cognition

CorC Correctability of Cognition

Bibliography

[1] L. Bainbridge. Ironies of automation. *Automatica*, 19(6):775–779, 1983.

[2] J. Brookshire, S. Singh, and R. Simmons. Preliminary results in sliding autonomy for assembly by coordination teams. In *Proceedings of the International Conference on Robots and Systems*, 2004.

[3] H. H. Clark and S. E. Brennan. *Perspectivies on socially shared cognition*, chapter Grounding in communication, pages 127–149. Washington, DC: American Psychological Association, 1991.

[4] J. W. Crandall and M. L. Cummings. Identifying predictive metrics for supervisory control of multiple robots. *IEEE Transactions on Robotics*, 23(5): 942–951, 2007.

[5] J. W. Crandall, M. A. Goodrich, D. R. Olsen, Jr., and C. W. Nielsen. Validating human-robot interaction schemes in multi-tasking environments. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 35(4):438–449, 2005.

[6] J.W. Crandall and M.A. Goodrich. Experiments in adjustable autonomy. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, volume 3, pages 1624 –1629, 2001.

[7] J. L. Drury, J. Scholtz, and H. A. Yanco. Awareness in human-robot interactions. In *Proceedings of the IEEE Conference on Systems, Man and Cybernetics*, pages 912–918, 2003.

[8] M. R. Endsley. Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors Society's 32nd Annual Meeting*, pages 97–101, 1988.

[9] M. R. Endsley, S. J. Selcon, T. D. Hardiman, and D. G. Croft. A comparative analysis of sagat and sart for evaluations of situation awareness. In *Proceedings of the Human Factors Society's 42nd Annual Meeting*, 1998.

[10] N. Giullian, D. Ricks, A. Atherton, M. Colton, M. Goodrich, and B. Brinton. Detailed requirements for robots in autism therapy. In *Proceedings of the IEEE International Conference on Systems Man and Cybernetics*, pages 2595 –2602, 2010.

[11] M. A. Goodrich and D. R. Olsen, Jr. Seven principles of efficient interaction. In *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, pages 3943–3948, 2003.

[12] Michael A. Goodrich and Alan C. Schultz. Human-robot interaction: a survey. *Foundations and Trends in Human-Computer Interaction*, 1(3):203–275, 2007.

[13] Marti A. Hearst. Trends & controversies: Mixed-initiative interaction. *IEEE Intelligent Systems*, pages 14–23, 1999.

[14] C. Heyer. Human-robot interaction and future industrial robotics applications. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4749–4754, 2010.

[15] G. Hoffman and C. Breazeal. Achieving fluency through perceptual-symbol practice in human-robot collaboration. In *Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction*, pages 1–8, 2008.

[16] T. Inagaki and K. Inoue. Adaptive choice of a safety management scheme upon an alarm under supervisory control of a large-complex system. *Reliability Engineering & System Safety*, 39(1):81 – 87, 1993.

[17] R. E. Kalman. On the general theory of control systems. In *Proceedings of the 1st International Congress of IFAC*, pages 481–492, 1960.

[18] R. E. Kalman. Mathematical description of linear dynamical systems. *SIAM Journal on Control and Optimization*, 1(2):152–192, 1963.

[19] S. Kiesler. Fostering common ground in human-robot interaction. In *Proceedings of the 14th IEEE International Workshop on Robots and Human Interactive Communication*, pages 792–734, 2005.

[20] H. Kozima, C. Nakagawa, and Y. Yasuda. Interactive robots for communication-care: a case-study in autism therapy. In *IEEE International Workshop on Robot and Human Interactive Communication*, pages 341 – 346, 2005.

[21] F. Mazzini, D. Kettler, J. Guerrero, and S. Dubowsky. Tactile robotic mapping of unknown surfaces, with application to oil wells. *IEEE Transactions on Instrumentation and Measurement*, 60(2):420–429, 2011.

[22] C. A. Miller, H. B. Funk, R. P. Goldman, J. Meisner, and P. Wu. Implications of adaptive vs. adaptable autonomy on decision making: Why "automated

adaptiveness" is not always the right answer. In *Proceedings of the 1st International Conference on Augmented Cognition*, 2005.

[23] Robin R. Murphy. *Introduction to AI Robotics*. MIT Press, Cambridge, MA, USA, 1st edition, 2000.

[24] D. Norman. *The Design of Everyday Things*. Basic Books, 1988.

[25] K. Ogata. *Modern Control Engineering*. Prentice-Hall, 5th edition, 2010.

[26] D. R. Olsen, Jr. and M. A. Goodrich. Metrics for evaluating human-robot interactions. In *Proceedings of the Performance Metrics for Intelligent Systems Workshop (PerMIS)*, 2003.

[27] R. Parasuraman, S. Galster, P. Squires, H. Furukawa, and C. Miller. A flexible delegation-type interface enhances system performance in human supervision of multiple robots: Empirical studies with RobotFlag. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 35:481–493, 2005.

[28] J. M. Riley, D. B. Kaber, and J. V. Draper. Situation awareness and attention allocation measures for quantifying telepresence experiences in teleoperation. *Hum. Factor. Ergon. Manuf.*, 14(1):51–67, 2004.

[29] B. Robins, K. Dautenhahn, R. te Boekhorst, A. Billard, S. Keates, J. Clarkson, P. Langdon, and P. Robinson. Effects of repeated exposure of a humanoid robot on children with autism. In *Designing a More Inclusive World*, pages 225–236. Springer-Verlag, 2004.

[30] B. Robins, K. Dautenhahn, and P. Dickerson. From isolation to communication: A case study evaluation of robot assisted play for children with autism with a minimally expressive humanoid robot. In *Proceedings of the 2009*

Second International Conferences on Advances in Computer-Human Interactions, pages 205–211, 2009.

[31] B. Scassellati. Affective prosody recognition for human-robot interaction. 2009.

[32] D. W. Schloerb. A quantitative measure of telepresence. *Presence: Teleoperators and Virtual Environments*, 4(1):64–80, 1995.

[33] B. Sellner, R. Simmons, and S. Singh. User modelling for principled sliding autonomy in human-robot teams. In *Multi-Robot Systems: From Swarms to Intelligent Automata*, 2005.

[34] T. B. Sheridan. *Telerobotics, Automation, and Supervisory Control*. MIT Press, 1992.

[35] T. B. Sheridan and W. L. Verplank. Human and computer control of undersea teleoperators. Technical report, MIT Man-Machine Systems Laboratory, 1978.

[36] SINTEF. Robotics for oil & gas platforms. *SINTEF Website*, Internet: http://www.sintef.no/home/Information-and-Communication-Technology-ICT/Applied-Cybernetics/Projects/Robotics-for-oil--gas-platforms, Feb. 9, 2000 [May 1, 2012].

[37] C. Stanton. Abu Dhabi signs nuclear power deal with South Korean group. *The National*, Internet: http://www.thenational.ae/news/uae-news/abu-dhabi-signs-nuclear-power-deal-with-south-korean-group, Dec. 28, 2009 [May 1, 2012].

[38] K. Stubbs, P. J. Hinds, and D. Wettergreen. Autonomy and common ground in human-robot interaction: A field study. *IEEE Intelligent Systems*, 22(2): 42–50, 2007.

[39] A. L. Thomaz and C. Breazeal. Teachable robots: Understanding human teaching behavior to build more effective robot learners. *Artificial Intelligence*, 172:716–737, 2008.