

Cognitive Telepresence in Human-Robot Interactions

Vahagn Harutyunyan, Vimitha Manohar, Issak Gezehei, and
Jacob W. Crandall

Masdar Institute of Science and Technology

Remote teleoperation of advanced, semi-autonomous robotic technologies has great potential in many industries critical to the economy and for the environment of the Middle East. Applications include maintenance of under-water oil wells, maintenance of nuclear power plants, counter-terrorism and national defense, law enforcement, remote sensing, and health care. In each of these applications, a user, likely without technology expertise, must operate a complex robot in uncertain and unknown environments. The nature of the tasks and environments encountered by the robot in these applications make it highly likely that the robot's limited autonomy will fail or be insufficient to complete the desired task. In this paper, we argue that, in such scenarios, cognitive telepresence, defined as the ability of the user to comprehend and control the robot's cognition, is an important design principle for human-robot systems. We compare and contrast cognitive telepresence to existing design principles commonly discussed in the literature, and define various metrics of cognitive telepresence. Finally, via two illustrative examples and a user study, we demonstrate the usefulness of cognitive telepresence as an important design principle of human-robot systems consisting of a user with limited technology expertise and a robot with limited and error-prone artificial intelligence.

Keywords: Human-robot interaction, supervisory control, teleoperation, metrics, system design, robot autonomy, user interfaces, cognitive telepresence

1. Introduction

As robotic systems mature, they will become more widely used in nations belonging to the Gulf Cooperation Council (GCC) in many different applications. First, the economies of many GCC countries are currently driven largely by the export of oil and gas, industries in which robotic systems are likely to play a prominent role, particularly in off-shore projects (e.g., Heyer, 2010; Mazzini, Kettler, Guerrero, & Dubowsky, 2011; SINTEF, 2012). Second, GCC nations are planning construction of nuclear power plants¹, wherein robots can fill important roles, particularly during emergencies (Parker & Draper, 1998). Third, safety concerns due to the current political climate

¹The Abu Dhabi Government recently announced in 2009 its plan to build four nuclear power plants to supply about a quarter of its power by the year 2020 (Stanton, 2009).

Authors retain copyright and grant the Journal of Human-Robot Interaction right of first publication with the work simultaneously licensed under a Creative Commons Attribution License that allows others to share the work with an acknowledgement of the work's authorship and initial publication in this journal.

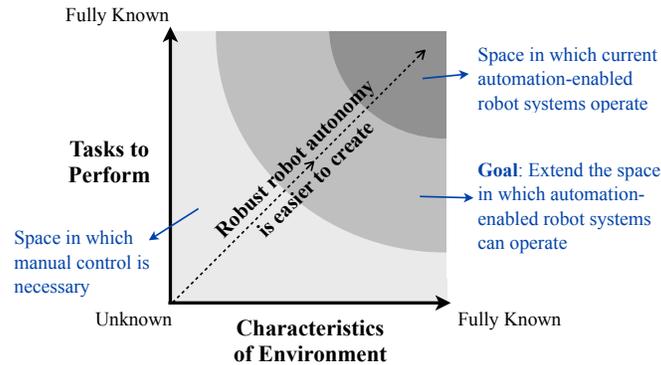


Figure 1. Currently, autonomy-enabled robot systems are possible when designers have high knowledge of the target task and environment. Appropriate enhancements to cognitive telepresence will make it possible for autonomy-enabled robot systems to operate with less prior knowledge.

surrounding GCC countries also encourage further robotics developments for national defense and counter-terrorism measures. Finally, as is the case throughout much of the world, the use of robots in health care (e.g. autism therapy (Giullian et al., 2010; Kozima, Nakagawa, & Yasuda, 2005; Robins, Dautenhahn, te Boekhorst, & Billard, 2004; Robins, Dautenhahn, & Dickerson, 2009; Scassellati, 2009)) is likely to become prominent in GCC countries.

While the attributes of future robotic systems for oil and gas, power plants, national defense, and autism therapy differ in many ways, these applications share three common characteristics. First, the robot's operational environment in each application will be somewhat unpredictable and complex, whether due to the unpredicted behavior of a child, clutter in the environment, unanticipated lighting conditions, etc. Thus, designers of robotic systems will not be able to fully envision the environments in which users will employ their robots *a priori*. Second, the tasks users will want their robots to perform may not be fully known to system designers *a priori*. If robots are capable, users will find uses for them that system designers did not envision. Third, the complexity of the robot required to performed effectively in each of these applications will be such that some robot autonomy, defined as the ability to carry out some function without human input (Parasuraman & Riley, 1997), will likely be necessary. We refer to such systems as *autonomy-enabled robot systems*, or systems in which robots autonomously perform the desired tasks with high proficiency and can often identify their own failures (Scheutz & Kramer, 2007).

The relationship between these three commonalities (knowledge of the environment, knowledge of tasks to perform, and proficiency of robot autonomy) is captured in Figure 1. When attributes of the environment and tasks are well-known, effective autonomy-enabled robot systems can be developed by system designers. However, in applications where system designers have only limited knowledge of the environment or tasks the robot will perform, developing robust robot autonomy is extremely difficult.

The typical solution for failed robot autonomy is to switch to manual teleoperation during detected autonomy failures. But manual teleoperation is often extremely difficult due to the complexity of the robot and difficulties with supplying the user with sufficient telepresence (Schloerb, 1995). Alternatively, robot systems can be created in which users can create their own robot autonomy via learning from demonstration (Argall, Chernova, Veloso, & Browning, 2009), scripting (Giullian et al., 2010), motion tracking (Atherton & Goodrich, 2011), or some other method. However, reliance

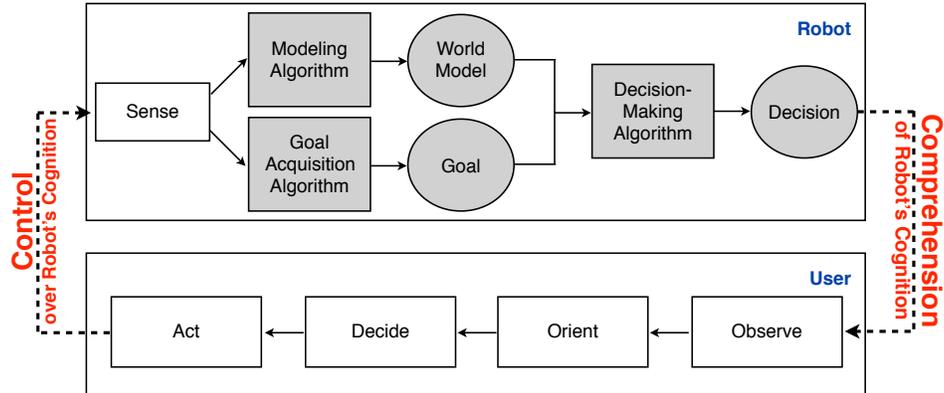


Figure 2. Cognitive telepresence refers to the user’s *comprehension* and *control* of the robot’s cognition (shaded shapes).

on such methods is (1) unlikely to guarantee robust robot autonomy, and (2) makes it difficult to design the system so that the robot reliably detects its own failures (Scheutz & Kramer, 2007). These unresolved challenges currently limit the use of autonomy-enabled robot systems.

In this paper, we advocate that autonomy-enabled robot systems be designed so they can operate effectively despite limitations and failures in robot autonomy. We argue that one critical design principle of such systems is *cognitive telepresence* (CT), which refers to a user’s ability to comprehend and control a robot’s cognition. CT focuses on overcoming failures of robot autonomy by regulating the robot’s cognition rather than overriding the robot’s actions manually.

The remainder of this paper proceeds as follows. In Section 2, we formally define and discuss CT. In Section 3, we compare and contrast CT to other popular and important design principles currently used in the development of human-robot systems, including situation awareness, levels of automation, adjustable autonomy, and neglect tolerance. We then discuss how CT should be measured in Section 4. In Section 5, we provide two examples to demonstrate how CT can be used as a design principle and to illustrate how it can be measured in specific human-robot systems. We also demonstrate the impact of CT on system effectiveness via a user study in Section 6. We discuss our conclusions in Section 7.

2. Cognitive Telepresence

When a robot has limited and error-prone autonomy, the user must (1) comprehend the robot’s cognition (i.e., the robot’s *artificial intelligence* or *autonomy*), and (2) modify the robot’s cognition when the user perceives that it is faulty or deficient. These two capabilities characterize CT.

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

CT is depicted in Figure 2, which shows simple cognitive models of both the user and the robot and interactions between them. The robot communicates a portion of its cognitive processes (boxes) and states (circles) to the user via the human-robot interface. This information helps the user to comprehend the robot’s cognition, and to determine whether the robot’s cognitive state and processes should be altered in any way. When necessary, the user provides input to the robot via the human-robot interface in attempt to correct elements of the robot’s cognition.

We refer to the two aspects of CT as *comprehension of the robot's cognition* (CompCo) and *control over the robot's cognition* (ConCo). We discuss each in turn.

2.1 Comprehension of the Robot's Cognition (CompCo)

CompCo refers to how much the user knows about the robot's current cognitive state and the processes (i.e., algorithms) it uses to derive its cognitive state. We now discuss the various elements of CompCo and the factors that contribute to it.

2.1.1 Elements of CompCo As shown in Figure 2, we break down the robot's cognitive state into three elements: its model of its environment (*world model*), its *goal*, and the actions it plans to take (i.e., its *decision*). Meanwhile, its corresponding cognitive processes consist of how 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 formulated by processing its inputs using its modeling algorithm. Thus, CompCo of the robot's world model refers to how well the user knows what the robot knows. For example, in a navigation task, the robot's world model may 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 robot's world model refers to the degree to which the user knows the robot's world model. 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 formed by its inputs. For example, the user might understand that the robot is not identifying a particular item correctly because the item does not meet a previously-defined color threshold in the video stream.

The robot's goal is formulated by processing its inputs using its goal-acquisition algorithm. Thus, CompCo of 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 with a marker on a map display indicating the robot's current destination. 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 of 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 the robot made the decision.

A particularly salient example of a mechanism designed to (partially) establish CompCo of the decision-making algorithm is provided in a study conducted by Thomaz and Breazeal (2008). In this study, a user was tasked with teaching a simulated robot how to bake a cake. The robot learned (using reinforcement learning) from the input the user provided while observing the robot act. In one particular scenario, the robot paused and glanced toward desirable actions when it was somewhat ambivalent between actions. This helped the user to understand that the robot placed similar values on actions, and that it did not know which action to take. This helped the user to comprehend how the robot makes decisions.

2.1.2 Contributing Factors To facilitate CompCo for each cognitive element, a human-robot interface must satisfy two necessary conditions: *observability* and *clarity*. As in the controls literature (Kalman, 1960, 1963), observability refers to the presence of information provided, in this case, to the user. For the user to possibly have full CompCo, enough information must be presented via the human-robot interface to make it possible for the user to fully model the robot's cognition.

However, observability alone does not guarantee CompCo. The user must also properly process

the information provided through the human-robot interface. Thus, the second interface requirement (clarity) refers to the ease at which the user can interpret information that the robot communicates. Returning to the example system created by Thomaz and Breazeal (2008), a human-robot interface could have displayed the robot's utility estimates for 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 uncertainty or ambivalence.

2.2 Control over the Robot's Cognition (ConCo)

Control over the robot's cognition (ConCo) refers to how well the user can control or change the robot's cognitive states and processes. We now discuss the elements and contributing factors of ConCo.

2.2.1 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 to alter each element as desired. If a user can quickly alter the robot's cognition as desired, then the user has high ConCo.

One straightforward way in which a user controls a robot's cognition is by overriding the robot's goals and decisions. For example, when a robot plans to traverse a path the user believes is undesirable, the user could be allowed to specify new waypoints or move existing waypoints to change the robot's intended path. However, such interactions can require substantial effort from users in the form of manual planning and implementation. Providing the user with the ability to modify the robot's world model (such as alerting the robot about an unseen hazard) can result in more simple and effective interactions that better utilize the robot's limited autonomy. Furthermore, when possible, allowing the user to make adjustments to the robot's modeling, goal-acquisition, or decision-making algorithms can provide long-term corrections to the robot's autonomy. We more thoroughly illustrate these concepts via examples and a user study in Sections 5 and 6.

2.2.2 Contributing Factors ConCo (over each cognitive element) is driven by two things: *controllability* and *expression*. Controllability, a term commonly used in the controls literature (Ogata, 2010), refers to the sovereignty the user has over each of the robot's cognitive elements. For full ConCo, it must be possible for the user to change the robot's cognition as desired via commands issued through the human-robot interface.

However, controllability alone does not guarantee ConCo. The second necessary factor of ConCo is expression (Olsen & Goodrich, 2003), which is similar in nature to the gulf of execution discussed by Norman (1988). Having determined what the robot's cognition should be, the user must be able to translate that intention into an input through the human-robot interface that will cause the desired outcome in the robot. Thus, expression refers to the user's ability to determine and communicate an input to the robot that has the desired effect on the robot's cognition.

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. To better quantify CT, we discuss its relationship with several of these principles, including situation awareness, telepresence, common ground, levels of automation, adjustable autonomy, and neglect tolerance.

3.1 Situation Awareness

A popular concept in human-robot systems is *situation awareness* (SA) or, more generally, *awareness*. Endsley defines SA as "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

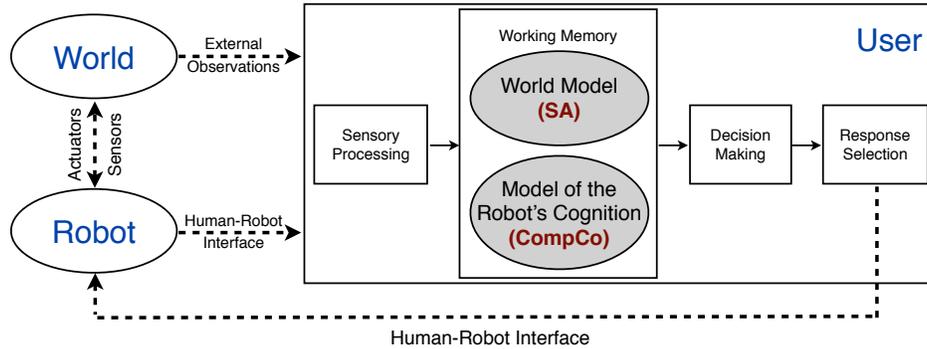


Figure 3. A human-robot interface should be designed so that a user can build an adequate model of the world (SA) and an adequate model of the robot's cognition (CompCo).

near future” (Endsley, 1988). For single-robot systems, Drury, Scholtz, and Yanco (2003) defined the term *HRI awareness* as “the understanding the human has of the location, activities, status, and surroundings of the robot.”

Our perception is that awareness in human-robot systems has typically focused on the user's understanding of the robot's environment, the robot's position in the environment, and the tasks the robot is carrying out, with some effort to begin modeling the user's awareness of the robot's status, such as the charge of the robot's battery (Drury et al., 2003). CompCo is related but more specific, as it focuses on the user's awareness of the robot's cognitive state and processes. Figure 3 illustrates the specific difference between SA and CompCo. SA is a comparison between the user's world model and the physical world, while CompCo is a comparison (in part) to the user's model of the robot's world model with the robot's actual world model. Thus, CompCo could potentially be re-phrased as *situation awareness of the robot's cognition*. However, CT also refers to the additional attribute of control over the robot's cognition, which SA does not.

Despite these distinctions, reasons for considering SA and CT are the same. The user must have both CT and SA to effectively interact with the robot so it accomplishes the desired task. Collectively, users' SA and CT define their ability to correct errors in the robot's autonomy. Effectively, the ability of the user to correct the robot's cognition, called *correctability of cognition* (or CorC), is a function of SA and CT. That is,

$$CorC = f(SA, CT). \quad (1)$$

We note an important relationship between SA and CT. A system that provides users with high ConCo but low SA gives users the ability to control the robot's cognition, but does not sufficiently help users know what the robot's cognition should be to adequately carry out the task. Likewise, high SA with low CT means that users will likely have to resort to manual control to overcome observed failings in the robot's autonomy.

3.2 Telepresence

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” (Schloerb, 1995). 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 (1992) argued that it is not

always desirable for a user to have full telepresence in human-robot systems. Likewise, too much CT can negatively impact the performance of a human-robot system.

3.3 Common Ground

Common ground is related in many ways to CompCo. It refers to the degree to which individuals share truthful mental models of each other (Clark & Brennan, 1991). Such mental models allow individuals to communicate effectively with each other with minimal effort (Kiesler, 2005). For example, human-robot interactions will be efficient if both the human and the robot have reasonable mental models of each other, which will in turn increase the predictability of the robot's behavior. Stubbs, Hinds, and Wettergreen (2007) concluded that causes of deficiencies in common ground between the user and the robot change as the robot's autonomy increases from insufficient contextual information and feedback to insufficient transparency. They claim that a robot should be able to adapt to the user's mental model in order to facilitate common ground. 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.

3.4 Levels of Automation

Another relevant and highly popular notion in human-machine systems is the concept of *levels of automation* (Sheridan & Verplank, 1978). A level of automation defines the relationship between a human and machine in performing a task. On one extreme, the human determines all the decisions and actions, while on the other extreme the machine decides and acts autonomously without any human input. Many intermediate levels of autonomy can also be achieved by combining the human's and the robot's decision-making in various ways.

We make two observations about the relationship between CT and levels of automation. First, in a sense, CT is exactly opposite to the concept of levels of automation. A level of automation refers to how the machine helps the user do a task, whereas CT addresses the problem of using humans to help a robot perform the task. Second, CT can be an important aspect at any level of automation short of manual control. Regardless of the role being played by the robot, the user may want or need to comprehend and control the robot's cognition. However, we argue that as systems increasingly rely on the robot's autonomy (i.e., higher levels of automation), CT becomes more important.

3.5 Adjustable Autonomy

One common method used to deal with failures in robot autonomy is adjustable autonomy (Sellner, Simmons, & Singh, 2005; Brookshire, Singh, & Simmons, 2004; Miller, Funk, Goldman, Meisner, & Wu, 2005; Parasuraman, Galster, Squires, Furukawa, & Miller, 2005). Adjustable autonomy refers to the process of switching (at the discretion of either the user or the robot) to a different level of automation depending on the current circumstances. 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 this difficult area before returning the responsibility of navigation to the robot.

However, manual control of a complex robot in an unknown and complex environment may require extreme time and effort from the user. A focus on CT offers an alternative. Rather than reverting to manual control (or some lower level of autonomy), the system can be designed so that the user can modify the robot's cognition, such as helping the robot identify passable or impassable regions in the world. Such an approach could potentially offer a more effective solution that requires less effort from the user.

Table 1: Explicit metrics of CT. Separate measures are taken for each cognitive element.

Name	Category	Definition	Description
Compre-hension	CompCo	$\frac{1}{\tau} \int_0^\tau E[X(t)] dt$	$E[X(t)]$ is the <i>CompCo accuracy curve</i> ; τ is a window of time over which the measure is taken
Control	ConCo	$\frac{1}{\tau} \int_0^\tau E[Y(t)] dt$	$E[Y(t)]$ is the <i>ConCo conformance curve</i> ; τ is a window of time over which the measure is taken

3.6 Neglect Tolerance: Neglect and Interaction Times

Neglect tolerance refers to the frequency and duration of human-robot interactions required for the robot to maintain acceptable performance for a given system (Crandall, Goodrich, Olsen, & Nielsen, 2005). In its simplest form, neglect tolerance is modeled by neglect time (the average amount of time the robot can be ignored by the operator before its performance drops below some threshold) and interaction time (the average duration of a single human-robot interaction) (Olsen & Goodrich, 2003; Goodrich & Olsen, 2003).

CT can substantially impact neglect tolerance. Appropriate levels of CT can help maximize the neglect tolerance of a system, while too little or too much CT will likely lower a system’s neglect tolerance. For example, consider a robot in a nuclear power plant where explosions have created a hazard the robot must avoid, but the robot does not recognize it as a hazard. One solution is to have the operator manually control the robot when it nears the hazard. In this case, neglect times would be low and interaction times could be high. Alternatively, if the operator were provided with the ability to alter the robot’s world model by communicating the position of the hazard in the environment, the robot could use this improved world model to navigate autonomously around the hazard, which would result in higher neglect tolerance than the first solution. Finally, the user could be allowed to interact with the robot’s modeling algorithm to teach it to recognize the new kind of hazard so it could also recognize similar hazards that may appear elsewhere. This third solution has the highest potential for increasing neglect times, though the interaction time required to facilitate this learning might be prohibitive. Thus, designers of human-robot systems must balance the potential risks and benefits of implementing high CT for each cognitive element.

4. Measuring Cognitive Telepresence

CT can be estimated explicitly via carefully controlled user studies. However, because CT is a mental process that we cannot fully observe and because CT interacts in complex ways with other cognitive processes (e.g., see Eq. 1), it is difficult to measure. Thus, partial or implicit measures are often more desirable.

Most measures of CT rely on a situation-specific metric, denoted $\mathcal{S}(m_1, m_2)$, which returns a value in the range $[0, 1]$. $\mathcal{S}(m_1, m_2)$ quantifies the *similarity* between cognitive models m_1 and m_2 . Higher values of $\mathcal{S}(m_1, m_2)$ indicate a closer match between m_1 and m_2 than lower outcomes. In the next section, we explicitly define $\mathcal{S}(m_1, m_2)$ for two example scenarios.

4.1 Explicit Metrics

CT can be explicitly measured in terms of CompCo and ConCo for each cognitive element. A summary of these two measures is given in Table 1. We describe each metric separately.

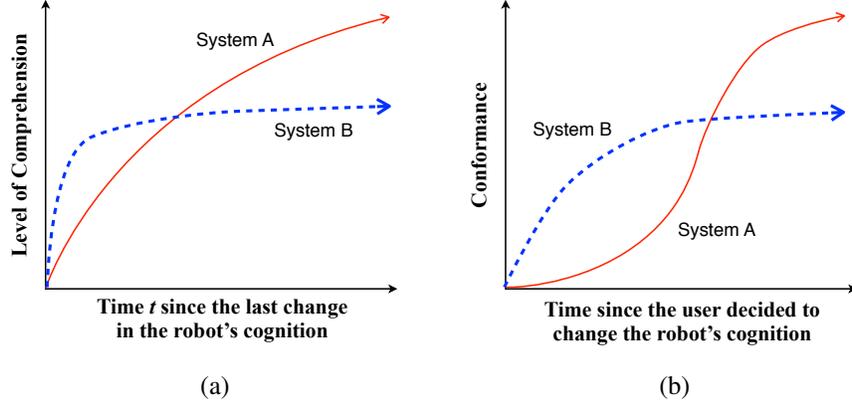


Figure 4. Hypothetical (a) CompCo accuracy and (b) ConCo conformance curves of two systems.

4.1.1 CompCo Comprehension of the robot's cognition is defined by a random process $X(t)$ that describes the extent to which users comprehend the robot's cognition (i.e., the *level of comprehension*) given the amount of time t since the last change in the robot's cognition. Formally, let m_t denote the robot's cognitive model at time t , and let \hat{m}_t denote the cognitive model that the user believes the robot has at time t . Then, $X(t) = \mathcal{S}(m_t, \hat{m}_t)$.

We call the mean function of $X(t)$, $E[X(t)]$, the *CompCo accuracy curve*. Figure 4a shows hypothetical CompCo accuracy curves of an arbitrary cognitive element for two systems. In System A, the user slowly gains CompCo until the robot's cognition is fully understood. On the other hand, the user more quickly gains (perhaps adequate) levels of comprehension with System B, but then struggles to obtain higher levels of comprehension thereafter. Depending on system characteristics and needs, either System A or B may be more desirable. Thus, Figure 4a illustrates the two important aspects of CompCo: speed of comprehension and peak comprehension. In general, effective systems provide the user with high CompCo more quickly than less effective systems.

Comprehension of a particular cognitive element is based on the CompCo accuracy curve of that element. Formally,

$$\text{Comprehension} = \frac{1}{\tau} \int_0^{\tau} E[X(t)] dt, \quad (2)$$

where τ is the window of time over which the user must respond to changes in the robot's cognition. τ is situation dependent. For high-workload and time-sensitive scenarios, τ should be smaller than in scenarios where the user is not required to respond as quickly.

The CompCo accuracy curve is difficult to measure due to its interdependence on other aspects of the system, such as the user's workload. However, it can be estimated via carefully controlled user studies in which users' levels of comprehension are tested at various time intervals.

4.1.2 ConCo There are two relevant attributes of ConCo: (1) the amount of time it takes the user to change the robot's cognition in the intended way (i.e., control speed), and (2) the extent to which the user can modify the robot's cognition (i.e., conformance). As with CompCo, these attributes are captured as part of a random process $Y(t) = \mathcal{S}(m_t, m_t^*)$, which defines over time (indexed from the time the user decides to change the robot's cognition) the *conformance* of the robot's cognitive model m_t with the model m_t^* the user intends it to have. We call $E[Y(t)]$ the *ConCo conformance curve*.

Table 2: Implicit metrics of CT. Separate measures are taken for each cognitive element.

Name	Category	Definition	Description
Observability	CompCo	$\mathcal{S}(m, \bar{m})$	Proportion of information in the robot's cognition that is observable via the human-robot interface
CompCo Assessment Technique	CompCo	% correct; time-to-answer	Proficiency (speed and accuracy) at answering questions related to the robot's current cognitive state and processes
Controllability	ConCo	% controllable	Proportion of information the user can change in the robot's cognition via the human-robot interface
Expression	ConCo	Eq. 4	Amount of interaction time required to effectuate the desired changes in the robot's cognition
CorC	CT	Eq. 5	Average correctness of the robot's cognition over time minus the baseline correctness \bar{C} (see text)

Figure 4b shows hypothetical ConCo conformance curves for two systems. The figure illustrates the potential trade-off between control speed and conformance.

Control for a given cognitive element is computed similarly to *Comprehension*. Formally,

$$Control = \frac{1}{\tau} \int_0^{\tau} E[Y(t)] dt, \quad (3)$$

where τ specifies the window of time over which the user must respond to changes in the robot's cognition. Like $E[X(t)]$, the ConCo conformance curve $E[Y(t)]$ can be estimated using carefully controlled user studies.

4.2 Partial Metrics

Given the difficulty of estimating *Comprehension* and *Control* via the CompCo accuracy and ConCo conformance curves, respectively, simpler methods for measuring CT might be more desirable. While not fully indicative, these measures can potentially allow system designers to more easily compare and contrast aspects of CT. Table 2 summarizes several partial metrics of CT, which we now describe in greater detail.

4.2.1 CompCo The CompCo of each cognitive element can be partially measured via two separate metrics. First, CompCo can be partially measured by the *Observability* of the cognitive element, which is the upper bound on the proportion of the robot's cognition that the user could possibly determine via the human-robot interface. Let m denote the robot's cognitive model and let \bar{m} denote the portion of the robot's cognitive model that the user could possibly derive from the human-robot interface. Then, $Observability = \mathcal{S}(m, \bar{m})$.

The second partial metric of CompCo that we propose is patterned after SAGAT (Endsley, 1988), a technique used to measure SA. The CompCo Assessment Technique involves periodically asking the users questions about the robot's cognition, and then assessing both the percentage of questions users are able to answer correctly and the amount of time it takes them to answer these questions. Questions can be administered by freezing the system and either blanking (Endsley, 1988) or not

blinking the screen (Endsley, Selcon, Hardiman, & Croft, 1998). The CompCo Assessment Technique only partially measures CompCo since it does not fully derive the CompCo accuracy curve, though it still produces both speed and accuracy measures.

4.2.2 ConCo The upper bound on the ConCo of each cognitive element of a system is its *Controllability* (Section 2.2.2), which represents the proportion of the robot’s cognition that the user can possibly modify via the human-robot interface. Let m denote the robot’s cognitive model and let m' denote the cognitive model that consists of modifiable aspects of the robot’s cognition. Then, $Controllability = \mathcal{S}(m, m')$.

A second partial metric of ConCo is *Expression*, which measures the user’s ability to determine and communicate the input that has the desired effect on the robot’s cognition. This metric is somewhat difficult to measure since the user’s “desired effect” can be difficult to deduce. However, expression can be estimated by observing how much effort the user must exert in order to change the robot’s cognition. We define expression as the change in the robot’s cognition per unit of interaction time. Let m_i denote the robot’s cognitive model in the absence of human-robot interactions, and let m_f denote the robot’s cognitive model given IT seconds of human-robot interactions. Then,

$$Expression = \frac{1 - \mathcal{S}(m_i, m_f)}{IT}. \quad (4)$$

4.2.3 Overall CT A useful metric for measuring overall CT is the correctability of cognition (CorC; see Eq. 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, for comparison among systems that produce similar SA, or systems in which users have near perfect SA, CorC is a useful measure of CT.

CorC is the average correctness of the robot’s cognition over some time T minus what the baseline average correctness of cognition the robot would have had if the user made no changes to the robot’s cognition (denoted \bar{C} , where $\bar{C} \in [0, 1]$). Formally, let m_t be the robot’s cognitive model at time t , and let m_t^r be the cognitive model that perfectly matches reality at time t . Then,

$$CorC = \frac{1}{T} \int_0^T \mathcal{S}(m_t, m_t^r) dt - \bar{C}. \quad (5)$$

The selection of the time interval T is typically made to include the full mission time. However, when the CorC of two systems is compared, the same T should be used to compute the CorC of both systems even though mission times may be different.

4.3 Implicit Metrics

The ultimate goal of any human-robot system is performance. Hence, system effectiveness itself is an important metric of CT. In many cases, we anticipate that performance is highly correlated with CT. A second implicit metric of CT is neglect tolerance, including neglect times and interaction times. Thus, the impact system characteristics have on neglect times and interactions times can be reflective of how those system characteristics affect CT. Sample implicit metrics of CT are summarized in Table 3.

5. Illustrative Examples

In this section, we use two example scenarios to demonstrate how CT can be used as a design principle and to illustrate how it can be measured for specific human-robot systems. In the first scenario, we develop and analyze a system for a combat scenario, namely detecting and disabling land mines. In the second study, we illustrate how CT can be used in the design of a robotic system for a free-play scenario in autism therapy.

Table 3: Implicit metrics of CT. Separate measures are taken for each cognitive element.

Name	Category	Definition	Description
System Effectiveness	CT	# of faults; Time-to-completion	A measure of the overall effectiveness of the system
Neglect Tolerance	CT	$\frac{\text{interaction time}}{\text{total time}}$	Ratio of interaction time to neglect time (or, similarly, overall mission time)

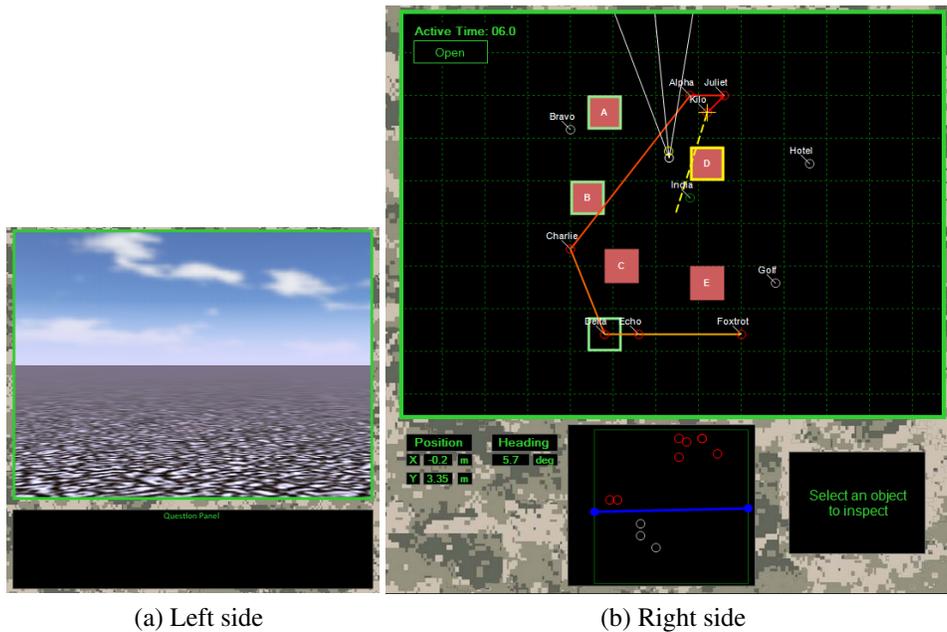


Figure 5. (a) Left and (b) right portions of the graphical user interface used to direct a robot in a mine patrol scenario.

5.1 Example I – Land Mine Patrol

In this scenario, a user operates a simulated remote robot to disable land mines.

5.1.1 Problem Description The robot patrols an area each day to dismantle land mines laid by the enemy during the night. The job of the robot each day is to identify and then disable all the mines in the area in as little time as possible. To do this, the robot must distinguish land mines from other objects (such as stones), and then navigate to and disable each one while avoiding contact with threat zones. A new set of land mines is placed in the area at the beginning of each day; the user is made aware of these threat zones via some external source (and not via the robot). Threat zones remain the same from day to day except that additional threat zones appear each day.

The user monitors and assists the robot in carrying out the mission via the graphical user interface depicted in Figure 5, which includes a sensor display (left hand side) and map display (right hand side). Since this robot operates primarily by vision, its sensor display consists of streamed video

Table 4: Descriptions of the robot’s autonomy. Shading indicates that the autonomy is error-prone.

Function	Element	Description
Object detection	World model	Detects and classifies objects as <i>stones</i> or <i>mines</i> based on color and shape. Detection of the location of objects is perfect, while classification accuracy is about 70–80%.
Threat detection	World model	Identifies threat zones in the environment. Detects just 10–15% of threat zones, and also occasionally identifies a safe zone as a threat zone.
Localization	World model	Flawlessly identifies its position in the environment.
Path planning	Decision-making	Calculates the shortest path from the robot’s current position through all detected mines.
Navigation	Decision-making	Navigates from source to destination while avoiding detected obstacles in the environment.

imagery. Via the map display, the user can observe the robot’s current position in relation to objects in the world (small circles with labels) and threat zones (pink rectangles). Additionally, the user can click on each object in the map display to view an image of that object (displayed in the bottom-right corner). This allows the user to determine which objects are mines and which are stones.

This scenario shares many common characteristics with other robotics applications pertinent to GCC countries (and elsewhere around the globe). In robotic systems for law enforcement, oil and gas, and maintenance of nuclear power plants, robots must detect and classify objects, navigate around obstacles, and repeatedly interact with a potentially dynamic environment.

5.1.2 Designing for Cognitive Telepresence System designers can follow a three-step process to help determine how to provide users with appropriate CT:

1. *Identify the robot’s autonomous functions.* Each day, the robot first scans the area for mines and stones. It then uses a classification algorithm to classify the objects from video images. After classifying the observed objects, the robot calculates the shortest path through all detected mines, and then navigates to and disables each one. Thus, in this system, the robot’s autonomy performs object and threat detection, localization, path planning, and navigation as summarized in Table 4.

2. *Define the robot’s cognitive model regarding aspects related to the robot’s error-prone autonomy.* In this system, the simulated robot performs localization, path planning, and navigation without error, but often fails to correctly detect objects and threats. Since object and threat detection correspond to the robot’s world model, we define the robot’s cognitive model as the set of items of which the robot is aware. Each item consists of a position and label (stone, mine, or threat). Two example cognitive models are shown in Table 5.

3. *Develop modifications to the system that allow users to comprehend and control the robot’s cognitive model.* Table 6 lists potential enhancements for this system.

5.1.3 Measuring Cognitive Telepresence Via user study, we can measure the CT of this system with respect to the partial metrics defined in Section 4. However, these metrics require that we define the similarity metric $\mathcal{S}(m_1, m_2)$, which compares cognitive models m_1 and m_2 . $\mathcal{S}(m_1, m_2)$ returns the fraction of items that are equivalent in m_1 and m_2 . Formally, let $O = m_1 \cup m_2$ be the

Table 5: Example world models for the system considered in the first example scenario.

ID	Location	Type	ID	Location	Type
1	(10, 7)	Mine	1	(10, 7)	Stone
2	(5, 6)	Stone	2	(5, 6)	Stone
3	(1, 10)	Stone	4	(3, 4)	Mine
4	(3, 4)	Mine	5	(5, 5)	Threat
5	(5, 5)	Threat	6	(5, 5)	Threat

Model m_1 Model m_2

Table 6: Potential modifications to the human-robot system to enhance cognitive telepresence.

Function	Category	Description
Show planned path	CompCo (decision)	Display the path the robot intends to follow.
Show robot's object labels	CompCo (world model)	Display the robot's classifications of each stone and mine in the map display. Objects classified as mines are displayed in red, stones are in gray.
Show robot's threat awareness	CompCo (world model)	Threats that the robot is aware of are outlined in green in the map display.
Change robot's object labels	ConCo (world model)	The user can change the robot's object labels by right-clicking on the object in the map display and selecting the option "switch type."
Communicate threats	ConCo (world model)	The user can inform the robot of an unseen threat by right-clicking on the threat in the map display.
Discard threats	ConCo (world model)	The user can tell the robot to discard a detected threat by right-clicking on the detected threat.
Change modeling algorithm	CT (modeling algorithm)	The user modifies the modeling algorithm by adjusting the linear threshold used to separate the classification of mines or stones in the algorithm's feature space (color and shape).

union of the items in world models m_1 and m_2 . Then,

$$\mathcal{S}(m_1, m_2) = \frac{\sum_{o \in O} I(m_1(o), m_2(o))}{|O|}, \quad (6)$$

where $I(a, b)$ is the indicator that returns 1 when item a is equivalent to item b and returns 0 otherwise, and where $m_i(o)$ represents m_i 's interpretation of item o . For the example models m_1 and m_2 shown in Table 5, m_1 and m_2 have equivalent models of items 2, 4, and 5. Thus, $\mathcal{S}(m_1, m_2) = \frac{3}{6} = 0.5$.

Given this similarity metric, we can measure observability, controllability, expression, and CorC (each as defined in Section 4) via user study. Additionally, the representation of the robot's world model indicates the kinds of questions that can be asked to measure CompCo via the CompCo Assessment Technique.

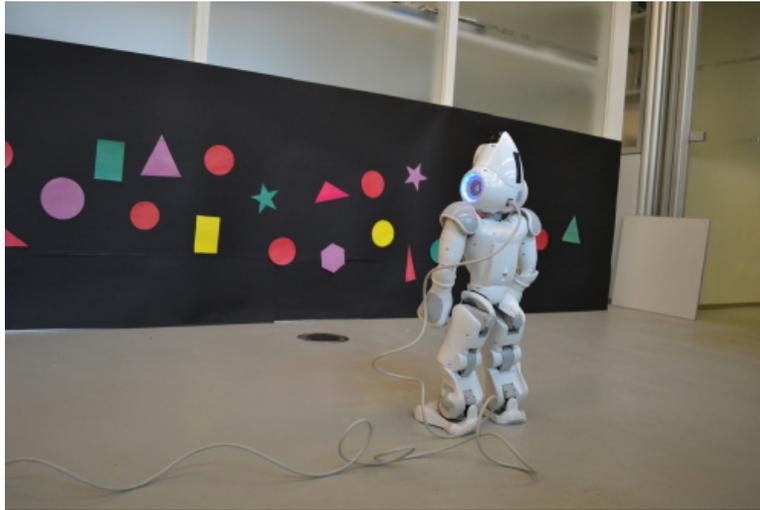


Figure 6. The robot must move to and count the shapes on the wall that match specific attributes.

5.2 Example II – Free Play in Autism Therapy

As a second example, we now discuss the design of a robotic system for another application important to GCC countries: using robots to assist in autism therapy for young children. This scenario has both similarities and differences with the previous scenario, and thus helps to identify how system designers can enhance and measure CT in human-robot systems.

5.2.1 Problem Description For some years now, robotic systems have been considered a potentially useful tool for use in therapy for children with autism (Giullian et al., 2010; Kozima et al., 2005; Robins et al., 2004, 2009; Scassellati, 2009). Despite much progress, the use of robotics in autism therapy is severely limited by the fact that therapists typically have trouble making robots work, and roboticists typically do not have a sound understanding of autism. We advocate that robot systems must allow therapists to easily configure and use robots with much less dependence on roboticists. CT is a critical component for providing this capability.

In our scenario, a humanoid Nao robot must interact with an autistic child and other caretakers (Giullian et al., 2010) in a discussion about images on a wall. For example, the child might be asked to count all of the red circles on the wall shown in Figure 6. To encourage interaction between the child and the caretakers, the robot also takes turns performing the task under critique of the autistic child.

This task can be difficult for the robot to perform without input from a user, particularly since the environment and the task might change over time. The robot could be controlled from another room by some collaborator (i.e., the user) (Giullian et al., 2010). However, given the large degrees of freedom of the robot and the difficulty with providing the user with high SA and telepresence, manual teleoperation of the robot would be quite difficult. It would be desirable if the robot could perform many tasks autonomously, though noise in the environment and potentially unexpected tasks provided in real-time are likely to cause this autonomy to periodically fail.

5.2.2 Designing for Cognitive Telepresence To design this system so that it provides users with adequate CT, system designers can follow the same three-step process identified in the first example

Table 7: Summary of the robot’s autonomy. Shading indicates that currently available robot autonomy for the function is error-prone.

Function	Element	Description
Determine objects to count	Goal	Determine the goal of the system by determining the features of images the robot should count. This is achieved by listening to the child and the caretaker.
Object detection	World model	Scan the wall to locate objects that match the goal description. Unanticipated lighting conditions and ill-defined object characteristics could cause failure.
Obstacle detection	World model	Detect hazards and obstacles the robot must avoid during navigation. Due to unknown and moving obstacles, this function could fail.
Localization	World model	Identify its position in the world using sensor data. Due to people (caretaker and child) moving about in the environment, this function is likely to occasionally fail.
Path planning	Decision-making	Calculate a path so that the robot moves before each detected target image on the wall.
Navigation	Decision-making	Navigate from source to destination while avoiding obstacles in the environment.

scenario:

1. *Identify the robot’s autonomous functions.* The autonomous functions performed by the robot in this task are listed in Table 7. The robot must first determine which objects to count. Once it has made this determination, it scans the wall to locate these objects. It then plans a path that takes it before each of the desired objects (so it can count them), and then navigates along this path while performing obstacle avoidance. Thus, as in the previous scenario, the robot performs object and obstacle detection, localization, path planning, and navigation. The robot must also deduce which images it should count in this example scenario, whereas this goal was hard-coded in the previous example.

2. *Define the robot’s cognitive model regarding aspects related to the robot’s error-prone autonomy.* Since this robotic system operates in the real world, more of its autonomy is subject to failure than the system described in the previous example scenario. This requires the robot’s cognitive model to be more complex, though some elements are the same. For this system, relevant aspects of a cognitive model include (1) the set of images the robot intends to count (represented by a location on the wall), (2) the set of obstacles the robot must avoid (defined by a bounding box), (3) the physical location and orientation of the robot (defined by the variables x , y , and θ), and (4) the features of the images on the wall the robot intends to count (size, shape, and color). Two example cognitive models are shown in Table 8.

3. *Develop modifications to the system that allow users to comprehend and control the robot’s cognitive model.* Table 9 lists seven potential enhancements for this system designed to improve CT.

5.3 Measuring Cognitive Telepresence

To measure CT, $\mathcal{S}(m_1, m_2)$ must be defined for each cognitive element in question. To measure CT with respect to both the robot’s goal and world model, we must define separate similarity metrics

Table 8: Example cognitive models for the system considered in the second example scenario.

<i>Description</i>	<i>Type</i>	<i>Parameters</i>
Image 1	World model	(1.5, 4.0, 0.3)
Image 2	World model	(3.3, 4.0, 0.4)
Image 3	World model	(0.1, 4.0, 0.5)
Obstacle	World model	(1.1, 1.6, 2.1, 1.9)
Robot location	World model	$x = 10.5$, $y = 20.1$, $\theta = 1.43$
Target Image Features	Goal	size=all, shape=circle, color=red

Model m_1

<i>Description</i>	<i>Type</i>	<i>Parameters</i>
Image 1	World model	(1.5, 4.0, 0.3)
Image 3	World model	(0.0, 4.0, 0.45)
Obstacle	World model	(1.1, 1.7, 2.0, 1.9)
Robot location	World model	$x = 10.7$, $y = 20.3$, $\theta = 1.6$
Target Image Features	Goal	size=all, shape=circle, color=blue

Model m_2

Table 9: Potential features of the system designed to provide appropriate cognitive telepresence.

Function	Category	Description
Display target images	CompCo (world model)	Provide a wall display which shows a mosaic of the wall (from video) overlaid with a depiction of the robot's beliefs about the locations of target images.
Change target images	ConCo (world model)	Give users the ability to add, delete, and alter the robot's beliefs about target images by manipulating the depiction of the robot's beliefs on the wall display.
Change modeling algorithm	CT (modeling algorithm)	Create a display that allows users to interactively adjust the properties (contrast, brightness, and color) of the video images used to detect target images.
Display position of robot and obstacles	CompCo (world model)	On a map of the robot's environment, display the robot's belief about (1) its own location and orientation and (2) the location of obstacles in the environment.
Change position of robot and obstacles	ConCo (world model)	Allow the user to alter the robot's belief about its own position and about obstacles (position and existence) by manipulating these objects in the map display.
Display features of target images	CompCo (Goal)	Display features of the images that the robot is trying to count to communicate the robot's goal to the user.
Change features of target images	ConCo (Goal)	Allow the user to change features that determine which images the robot should count. The user can select features of objects the robot should count from drop-down menus.

for the robot’s goal ($\mathcal{S}_{\text{goal}}(m_1, m_2)$) and world model ($\mathcal{S}_{\text{model}}(m_1, m_2)$). This metric computes the fraction of matching goal features in cognitive models m_1 and m_2 . Formally,

$$\mathcal{S}_{\text{goal}}(m_1, m_2) = \frac{\sum_{j \in \{\text{size, shape, color}\}} I(m_1^j, m_2^j)}{3}, \quad (7)$$

where m_i^j is m_i ’s interpretation of feature j and $I(a, b)$ is, as in Eq. 6, the indicator function. For the two example cognitive models shown in Table 8, $\mathcal{S}_{\text{goal}}(m_1, m_2) = \frac{2}{3}$, since the size and shape match, but the color does not.

$\mathcal{S}_{\text{model}}(m_1, m_2)$ is defined similarly to Eq. 6 with two exceptions. First, robot location is considered an item in the cognitive model together with obstacles and target images. Second, unlike the first scenario, Nao does not always correctly identify the location of target images, obstacles, and its own position; it has some errors. Thus, rather than using the indicator function as a measure of comparison between the interpretations of items made by two models, we use the L_2 -norm, capped by some max distance (we use $d_{\text{max}} = 0.5$ meters). Formally, let $L(o_1, o_2) = \min(\|o_1 - o_2\|_2, d_{\text{max}})$ be the distance between items o_1 and o_2 . Then,

$$\mathcal{S}_{\text{world}}(m_1, m_2) = \frac{\sum_{o \in O} (d_{\text{max}} - L(m_1(o), m_2(o)))}{|O| \cdot d_{\text{max}}}, \quad (8)$$

where the set O is the set of all world-model-related items contained in both m_1 and m_2 , and $m_i(o)$ is m_i ’s interpretation of item o . We could potentially weigh the items according to their importance. However, such importance is unclear to us, so we weighed them equally.

For the two example cognitive models shown in Table 8, Eq. 8 produces

$$\mathcal{S}_{\text{world}}(m_1, m_2) = \frac{0.5 + 0.0 + 0.45 + 0.359 + 0.17}{5 \cdot 0.5} = \frac{1.479}{2.5} = 0.592.$$

6. User Study

To demonstrate the effects of CT on the performance characteristics of the system, we conducted a user study to compare and contrast five different variations in the system discussed in example I.

6.1 Experimental Design

In this user study, we evaluated the five systems outlined in Table 10. The systems differ with respect to the seven potential methods for enhancing CT listed in Table 6. All the systems utilize the same robot autonomy, and differ only with respect to the features intended to enhance CT. System 0 provides users with minimal comprehension and control over the robot’s cognition. System 1 potentially provides users with some CompCo over the robot’s world model and decision. System 2 is identical to System 1 except that it allows users to alter the robot’s object labels and to notify the robot of threats. System 3 also gives users the ability to discard threats. It also communicates the robot’s knowledge of threats to users. Finally, System 4 allows users to alter the robot’s classification algorithm by adjusting the linear threshold used to separate the classification of mines or stones in the algorithm’s feature space (color and shape). In this system, users are shown locations of the objects as well as the classification threshold in feature space on the display located at the bottom-middle of Figure 5b. Users can adjust the linear classification threshold by moving the endpoints of the threshold. Thus, while System 4 does not improve the controllability of the robot’s world model relative to System 3, it does give users some control over the robot’s modeling algorithm.

Twenty individuals from the Masdar Institute community between the ages of 23 and 32 (mean age was 26.6) participated in the study. Each participant was assigned one of the five systems. After

Table 10: Attributes of five systems evaluated in the user study.

<i>Feature</i>	<i>System 0</i>	<i>System 1</i>	<i>System 2</i>	<i>System 3</i>	<i>System 4</i>
Show planned path		✓	✓	✓	✓
Show robot's object labels		✓	✓	✓	✓
Show robot's awareness of threats				✓	✓
Change robot's object labels			✓	✓	✓
Communicate threats			✓	✓	✓
Discard threats				✓	✓
Change modeling algorithm					✓

being trained on the assigned system until they felt comfortable using it, participants used the system to deactivate land mines in a test scenario consisting of three simulated days. At 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 also placed in the environment, of which the robot was initially aware of only one. These three threats persisted in the environment from Day 1 through Day 3. Two additional threats (initially unknown to the robot) were placed in the environment in Day 2. In Day 3, the robot detected a safe zone around one of the mines as a threat zone. The robot remembered the location of known threats from one day to the next.

To assess relative SA and CompCo (via the CompCo Assessment Technique), participants were asked questions designed to assess their understanding of the environment and the robot's cognitive state. A new question appeared 23 seconds after the previous question was answered. When a question was displayed, the system was paused until the user answered it. All displays were available to participants 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?" The possible answers to these questions were: (1) Rock, (2) Mine, and (3) Unsure. Similar questions testing participants' knowledge of the robot's knowledge of threats and object classification were also asked. For example, participants were asked: "What does the robot think object Alpha is?"

Reported statistical tests are results of two-way mixed analysis of variance, with *system* as a between-subjects factor with five levels (Systems 1–5) and *day* as a within-subjects factor with three levels (Days 1–3).

6.2 Results

The purpose of this user study was (1) to demonstrate how CT can impact the performance of human-robot systems and (2) to demonstrate the effectiveness of the various metrics of CT. We begin by summarizing the overall performance of the five systems, which is shown in Figure 7. For each system, the figures show the average amount of time the robot spent in threat zones per day (Figure 7a), the average number of mines that were not deactivated per day (Figure 7b), and the average time-to-completion per day (Figure 7c).

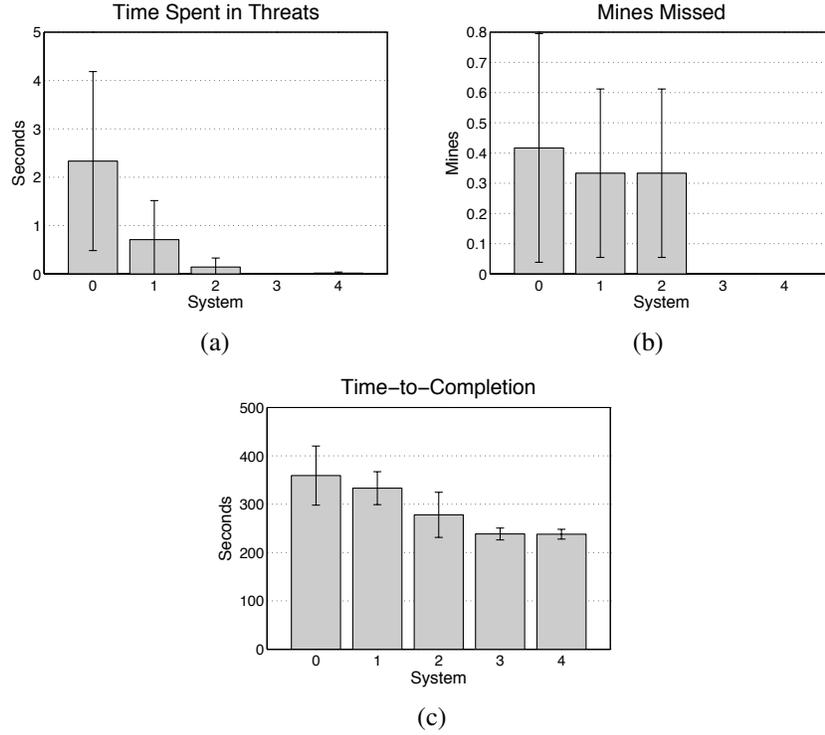


Figure 7. System effectiveness averaged over all days in terms of (a) time spent in threats, (b) mines missed, and (c) time-to-completion. Error bars show a 95% confidence interval on the mean.

For each of these performance metrics, higher-numbered systems tended to have higher performance than lower-numbered systems. A two-way mixed analysis of variance showed that time-to-completion varied significantly across systems ($F(4, 15) = 13.59, p < 0.001$). Pairwise comparisons show that time-to-completion was statistically different with Systems 0 and 1 than with both System 3 ($p = 0.001$ and $p = 0.002$, respectively) and System 4 ($p = 0.001$ and $p = 0.002$, respectively).

To help understand why performance was better in higher-numbered systems, we analyzed the CT of the systems using the partial metrics of Observability, Controllability, Expression, CorC, and measures derived from the CompCo Assessment Technique. In our study, the average Observability and Controllability of the systems (based on the similarity metric given in Eq. 6), are shown in Table 11. Both Systems 3 and 4 provide full Observability and Controllability over the robot's world model.

Table 11: Average Observability and Controllability of the five systems over all three days.

	System 0	System 1	System 2	System 3	System 4
Observability	0.00	0.70	0.70	1.00	1.00
Controllability	0.00	0.00	0.98	1.00	1.00

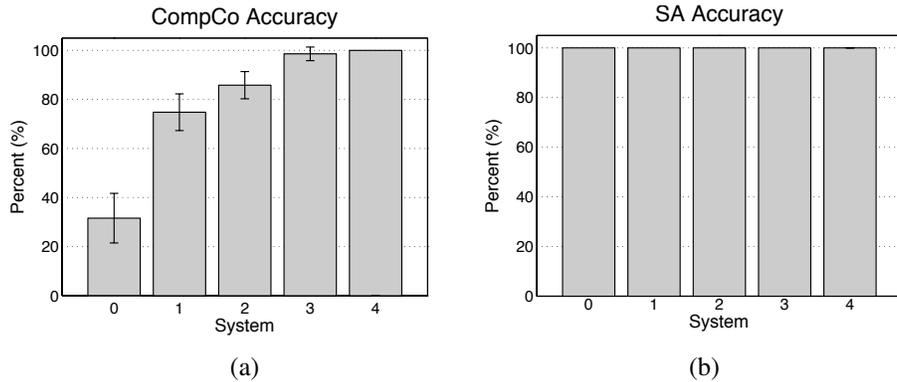


Figure 8. Percentage of CompCo and SA questions answered correctly. Error bars show a 95% confidence interval on the mean.

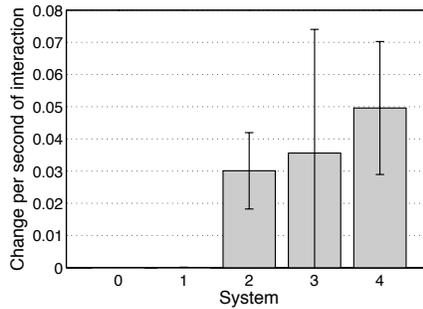


Figure 9. Expression for each system. Error bars show a 95% confidence interval on the mean.

Observability appears to have translated directly into CompCo. A two-way mixed analysis of variance showed that CompCo accuracy (Figure 8a), i.e. the percentage of questions answered correctly when administering the CompCo Assessment Technique, varied substantially and significantly across systems ($F(4, 15) = 78.05, p < 0.001$). Pairwise comparisons showed that CompCo accuracy was statistically different with Systems 3 and 4 than with Systems 0 and 1 ($p \leq 0.001$ in each case). While subjects had nearly perfect CompCo accuracy when using Systems 3 and 4, they had very low CompCo accuracy when using System 0. In contrast, participants always correctly answered the SA-related questions in each system (Figure 8b). The amount of time it took subjects to answer both SA- and CompCo-related questions did not vary substantially across systems.

Measures of Expression and CorC, which measure how well users are able to change the robot’s cognition, produced similar trends as Observability, Controllability, and CompCo accuracy (Figures 9 and 10). As expected, the average Expression and CorC were zero for both Systems 0 and 1, as these systems did not allow users to modify the robot’s cognition. Average Expression and CorC were both highest in System 4. Since CorC is a function of both SA and CT, and since SA did not appear to vary across systems, much of the difference among systems with respect to CorC appears to be due to CT. Both the measures of expression and CorC demonstrated that users were better able to comprehend and control the robot’s cognition using System 4 than when using the other systems.

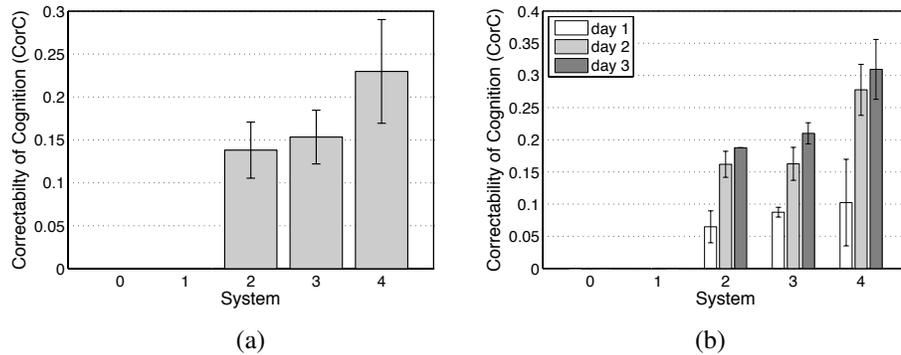


Figure 10. CoC for each system. Error bars show a 95% confidence interval on the mean.

Recall that System 4 had all of the capabilities of the other systems. It also provided participants with the ability to alter the robot’s modeling algorithm by adjusting the classification threshold. Participants typically adjusted this threshold within the first day so that the robot always correctly distinguished mines from stones. This produced substantially higher measures of CoC for Days 2 and 3 (Figure 10b) when no additional interaction was required in this regard. Thus, while both System 3 and 4 had comparable Controllability, participants were able to control the robot’s cognition the best when using System 4. Furthermore, changes made to the robot’s modeling algorithm had longer-lasting effects than did direct changes to the robot’s world model.

In our study, the average system effectiveness was higher for higher-numbered systems than for lower-numbered systems. Likewise, each of the partial measures of CT shows that CT was higher, on average, for higher-numbered systems than for lower-numbered systems. Given that the systems only differed by features designed for increasing comprehension and control over the robot’s world model and decisions, this correlation supports the argument that CT is an important design principle for autonomy-enabled human-robot systems. Providing users with the ability to observe and modify the robot’s cognition can be an effective method for overcoming deficiencies in the robot’s autonomy.

7. Conclusions

As robotic systems mature, they are likely to come into high demand in GCC countries. Applications include robots for the oil and gas industries, national defense and counter-terrorism, power plant maintenance, and health care. Each of these applications is likely to share a set of common characteristics. First, the environments in which the robot operates in each application are noisy, dynamic, and unknown. Second, the tasks that users will want their robots to perform may not be fully known to system designers *a priori*. Third, the complexity of the robot will be such that some robot autonomy will be necessary to perform the desired tasks proficiently. Unfortunately, given the uncertainty in both the task and environment, 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 paper, we argued that cognitive telepresence (CT) (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 defined and discussed this design principle in detail, and also discussed how it complements common design principles already defined and used in the literature. We then presented two illustrative example scenarios in which we (1) illustrated how CT can be used as a

design principle and (2) demonstrated how it can be measured in two different scenarios. We then showed that CT can have an important impact on the overall performance of human-robot systems via a user study.

Despite continued improvements, robot autonomy will continue to be error-prone and limited in the foreseeable future for many unstructured environments. This will continue to limit the use of robots unless systems can be designed in which operators can effectively manage deficient robot autonomy. Our results show that CT is an important attribute of such systems.

8. Acknowledgments

This work was supported by the Government of Abu Dhabi, United Arab Emirates.

References

- Argall, B. D., Chernova, S., Veloso, M., & Browning, B. (2009). A survey of robot learning from demonstration. *Robotics and Autonomous Systems*, 57, 469–483. <http://dx.doi.org/10.1016/j.robot.2008.10.024>.
- Atherton, J. A., & Goodrich, M. A. (2011). Supporting clinicians in robot-assisted therapy for autism spectrum disorder: Creating and editing robot animations with full-body motion tracking. In *Human-Robot Interaction: Perspectives and Contributions to Robotics from the Human Sciences Workshop at Robotics Science and Systems*. Los Angeles, CA, USA.
- Brookshire, J., Singh, S., & Simmons, R. (2004). Preliminary results in sliding autonomy for assembly by coordination teams. In *Proceedings of the International Conference on Robots and Systems* (p. 706-711). <http://dx.doi.org/10.1109/IROS.2004.1389435>.
- Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on Socially Shared Cognition* (Vol. xiii, pp. 127–149). Washington, DC, USA: American Psychological Association. <http://dx.doi.org/10.1037/10096-006>.
- Crandall, J. W., Goodrich, M. A., Olsen, D. R., Jr., & Nielsen, C. W. (2005). 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. <http://dx.doi.org/10.1109/TSMCA.2005.850587>.
- Drury, J. L., Scholtz, J., & Yanco, H. A. (2003). Awareness in human-robot interactions. In *Proceedings of the IEEE Conference on Systems, Man and Cybernetics* (pp. 912–918). <http://dx.doi.org/10.1109/ICSMC.2003.1243931>.
- Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors Society's 32nd Annual Meeting* (pp. 97–101). Anaheim, CA, USA.
- Endsley, M. R., Selcon, S. J., Hardiman, T. D., & Croft, D. G. (1998). A comparative analysis of SAGAT and SART for evaluations of situation awareness. In *Proceedings of the Human Factors Society's 42nd Annual Meeting*. <http://dx.doi.org/10.1177/154193129804200119>.
- Giullian, N., Ricks, D., Atherton, A., Colton, M., Goodrich, M., & Brinton, B. (2010). Detailed requirements for robots in autism therapy. In *Proceedings of the IEEE International Conference on Systems Man and Cybernetics* (p. 2595 -2602). <http://dx.doi.org/10.1109/ICSMC.2010.5641908>.
- Goodrich, M. A., & Olsen, D. R., Jr. (2003). Seven principles of efficient interaction. In *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics* (pp. 3943–3948). <http://dx.doi.org/10.1109/ICSMC.2003.1244504>.
- Heyer, C. (2010). Human-robot interaction and future industrial robotics applications. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 4749–4754). <http://dx.doi.org/10.1109/IROS.2010.5651294>.
- Kalman, R. E. (1960). On the general theory of control systems. In *Proceedings of the 1st International Congress of International Federation of Automatic Control* (pp. 481–492). Moscow, USSR.
- Kalman, R. E. (1963). Mathematical description of linear dynamical systems. *SIAM Journal on Control and Optimization*, 1(2), 152-192. <http://dx.doi.org/10.1137/0301010>.
- Kiesler, S. (2005). Fostering common ground in human-robot interaction. In *Proceedings of the 14th IEEE International Workshop on Robots and Human Interactive Communication* (pp. 792–734). <http://dx.doi.org/10.1109/ROMAN.2005.1513866>.

- Kozima, H., Nakagawa, C., & Yasuda, Y. (2005). Interactive robots for communication-care: A case-study in autism therapy. In *IEEE International Workshop on Robot and Human Interactive Communication* (p. 341 - 346). <http://dx.doi.org/10.1109/ROMAN.2005.1513802>.
- Mazzini, F., Kettler, D., Guerrero, J., & Dubowsky, S. (2011). Tactile robotic mapping of unknown surfaces, with application to oil wells. *IEEE Transactions on Instrumentation and Measurement*, 60(2), 420–429. <http://dx.doi.org/10.1109/TIM.2010.2084270>.
- Miller, C. A., Funk, H. B., Goldman, R. P., Meisner, J., & Wu, P. (2005). 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*. Las Vegas, Nevada, USA.
- Norman, D. (1988). *The design of everyday things*. New York: Basic Books.
- Ogata, K. (2010). *Modern control engineering* (5th ed.). Prentice-Hall. <http://dx.doi.org/10.1115/1.3426465>.
- Olsen, D. R., Jr., & Goodrich, M. A. (2003). Metrics for evaluating human-robot interactions. In *Proceedings of the Performance Metrics for Intelligent Systems Workshop (PerMIS)*. Gaithersburg, MD, USA.
- Parasuraman, R., Galster, S., Squires, P., Furukawa, H., & Miller, C. (2005). 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. <http://dx.doi.org/10.1109/TSMCA.2005.850598>.
- Parasuraman, R., & Riley, V. A. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230–253. <http://dx.doi.org/10.1518/001872097778543886>.
- Parker, L. E., & Draper, J. V. (1998). Robotics applications in maintenance and repair. In S. Y. Nof (Ed.), *Handbook of Industrial Robotics* (2nd ed., Vol. 1, pp. 1023–1036). Hoboken, NJ: John Wiley & Sons, Inc.
- Robins, B., Dautenhahn, K., & Dickerson, P. (2009). 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 2nd International Conferences on Advances in Computer-Human Interactions* (pp. 205–211). <http://dx.doi.org/10.1109/ACHI.2009.32>.
- Robins, B., Dautenhahn, K., te Boekhorst, R., & Billard, A. (2004). Effects of repeated exposure of a humanoid robot on children with autism. In S. Keates, J. Clarkson, P. Langdon, & P. Robinson (Eds.), *Designing a More Inclusive World* (p. 225-236). London: Springer Verlag. http://dx.doi.org/10.1007/978-0-85729-372-5_23.
- Scassellati, B. (2009). Affective prosody recognition for human-robot interaction. In *Microsoft Research's External Research Symposium*. Redmond, WA, USA.
- Scheutz, M., & Kramer, J. (2007). Reflection and reasoning mechanisms for failure detection and recovery in a distributed robotic architecture for complex robots. In *Proceedings of the IEEE International Conference on Robotics and Automation* (pp. 3699–3704). <http://dx.doi.org/10.1109/ROBOT.2007.364045>.
- Schloerb, D. W. (1995). A quantitative measure of telepresence. *Presence: Teleoperators and Virtual Environments*, 4(1), 64–80.
- Sellner, B., Simmons, R., & Singh, S. (2005). User modelling for principled sliding autonomy in human-robot teams. In *Multi-Robot Systems: From Swarms to Intelligent Automata* (p. 197-208). http://dx.doi.org/10.1007/1-4020-3389-3_16.
- Sheridan, T. B. (1992). *Telerobotics, automation, and supervisory control*. Cambridge, MA: MIT Press.
- Sheridan, T. B., & Verplank, W. L. (1978). *Human and computer control of undersea teleoperators* (Tech. Rep.). Man-Machine Systems Laboratory, Massachusetts Institute of Technology, Cambridge, MA.
- SINTEF. (2012). *Off-shore robotics – Robotics for oil & gas platforms*. SINTEF. Retrieved from <http://www.sintef.no/home/Information-and-Communication-Technology-ICT/Applied-Cybernetics/Projects/Robotics-for-oil--gas-platforms>.
- Stanton, C. (2009). *Abu Dhabi signs nuclear power deal with South Korean group*. The National. Retrieved from <http://www.thenational.ae/news/uae-news/abu-dhabi-signs-nuclear-power-deal-with-south-korean-group>.
- Stubbs, K., Hinds, P. J., & Wettergreen, D. (2007). Autonomy and common ground in human-robot interaction: A field study. *IEEE Intelligent Systems*, 22(2), 42–50. <http://dx.doi.org/10.1109/MIS.2007.21>.
- Thomaz, A. L., & Breazeal, C. (2008). Teachable robots: Understanding human teach-

ing behavior to build more effective robot learners. *Artificial Intelligence*, 172, 716–737.
<http://dx.doi.org/10.1016/j.artint.2007.09.009>.

Authors' names and contact information: Vahagn Harutyunyan, vharutyunyan@masdar.ac.ae; Vimitha Manohar, vmanohar@masdar.ac.ae; Issak Gezehei, igezehei@masdar.ac.ae; Jacob W. Crandall, jcrandall@masdar.ac.ae; Computing and Information Science Program, Masdar Institute of Science and Technology, Abu Dhabi, UAE.