Toward a Framework for Levels of Robot Autonomy in Human-Robot Interaction

Jenay Beer
University of South Carolina - Columbia, jbeer@cse.sc.edu

A. D. Fisk

W. A. Rogers

Publication Info
Published in *Journal of Human-Robot Interaction*, Volume 3, Issue 2, 2014, pages 74-99.

Toward a Framework for Levels of Robot Autonomy in Human-Robot Interaction

Jenay M. Beer
College of Engineering and Computing and College of Social Work, University of South Carolina
Arthur D. Fisk
School of Psychology, Georgia Institute of Technology
and
Wendy A. Rogers
School of Psychology, Georgia Institute of Technology

Autonomy is a critical construct related to human-robot interaction (HRI) and varies widely across robot platforms. Levels of robot autonomy (LORA), ranging from teleoperation to fully autonomous systems, influence the way in which humans and robots interact with one another. Thus, there is a need to understand HRI by identifying variables that influence—and are influenced by—robot autonomy. Our overarching goal is to develop a framework for LORA in HRI. To reach this goal, our framework draws links between HRI and human-automation interaction, a field with a long history of studying and understanding human-related variables. The construct of autonomy is reviewed and redefined within the context of HRI. Additionally, this framework proposes a process for determining a robot's autonomy level by categorizing autonomy along a 10-point taxonomy. The framework is intended to be treated as a guideline for determining autonomy, categorizing the LORA along a qualitative taxonomy and considering HRI variables (e.g., acceptance, situation awareness, reliability) that may be influenced by the LORA.

Keywords: human-robot interaction, automation, autonomy, levels of robot autonomy, framework

1. Introduction

Autonomy: from Greek *autos* ("self,") and *nomos* ("law")
"*I am putting myself to the fullest possible use…*" –HAL 9000 (2001: Space Odyssey)

A focus on robot autonomy has scientific importance beyond the pop culture goal of creating a machine that demonstrates some level of artificial free will. Robot autonomy broadly refers to the system's capability to carry out its own processes and operations. This is of particular importance within the field of human-robot interaction (HRI) because a robot's autonomy will impact the tasks it is able to perform, the level and frequency of interaction with human operators, and the reliability with which that robot can perform in an environment.

Determining appropriate autonomy in a machine (robotic or otherwise) is not an exact science. An important question is not, "What can a robot do?" but rather, "What should a robot do,

Journal of Human-Robot Interaction, Vol. 3, No. 2, 2014, Pages 74-99, DOI 10.5898/JHRI.3.2.Beer

and to what extent?" A scientific base of empirical research can guide designers in identifying appropriate tradeoffs to determine which functions and tasks to allocate to either a human or a robot. Autonomy is a central factor determining the effectiveness of the human-machine system. Therefore, understanding robot autonomy is essential to understand HRI.

Autonomy has been conceptualized in different fields in different ways. In human-automation interaction, for example, autonomy has been largely explained as an allocation of function between a human and a robot. In HRI, there are two schools of thought in conceptualizing autonomy: (1) higher robot autonomy requires less frequent interaction; and (2) higher robot autonomy requires higher levels or more sophisticated forms of interaction. The disparate way autonomy has been conceptualized in HRI is due, in part, to the multidisciplinary nature of the field but also to the lack of an integrated autonomy framework. To parse the complexity of autonomy, a framework of robot autonomy is needed. Development of a framework on levels of autonomy for human-robot interaction not only holds promise to conceptualize and better understand the construct of autonomy, but also to account for human cognitive and behavioral responses (e.g., situation awareness, workload, acceptance) within the context of HRI.

This proposed framework focuses on service robots. Although this category is broad, certain shared characteristics are relevant to autonomy and HRI. First, service robots of varying degrees of autonomy have been applied to a range of applications, such as domestic assistance, healthcare nursing tasks, search and rescue, and education. Second, due to the range of service applications, human-robot interaction will often be necessary, and service robots of varying autonomy levels may be expected to interact with humans having limited or no formal training (Thrun, 2004).

2. Goals and Contributions

Although previous models and frameworks have addressed autonomy in HRI (Feil-Seifer, Skinner, & Mataric, 2007; Goodrich & Olsen, 2003; Goodrich & Schultz, 2007; Huang, Pavek, Albus, & Messina, 2005; Thrun, 2004; Yanco & Drury, 2004*)* and automation (Endsley & Kaber, 1999; Parasuraman, Sheridan, & Wickens, 2000; Sheridan & Verplank, 1978*)*, they have not provided a cohesive definition and conceptualization of robot autonomy that allows designers and researchers to identify a robot's autonomy level and consider how it might impact the interaction between the robot and the human. The overarching goal of developing a comprehensive framework of autonomy was to expand upon the current models and consider, in depth, the role of autonomy in HRI. To this end, an in-depth literature review revealed several necessary elements for this framework:

First, a new definition of autonomy was needed to clarify how this concept should be considered for HRI. The first element of this framework is a definition that is specific to robotics and HRI through the integration of various definitions that have been used in HRI and related fields

Second, autonomy literature was reviewed to develop a coherent understanding of autonomy in HRI. Here, elements from research on automation, as well as considerations from research on HRI, were synthesized into a set of guidelines meant to provide guidance for designers in identifying the appropriate autonomy level of a robot. Designers and researchers can use these guidelines to consider what level of autonomy is appropriate for their robot and the impact of autonomy on how the human and robot interact.

Finally, within these guidelines, a taxonomy of robot autonomy levels was proposed. We defined this taxonomy as a representation of a classification technique onto a set of ordered categories, which can be seen as lying along a continuum.

All of these elements: (1) definition, (2) guidelines, and (3) taxonomy make up the framework. The framework we have proposed is broad by nature. Our aim was to consider autonomy for service robots, which may encompass a wide range of robot types, capabilities, and operational environments. The purpose of this framework is to take a conceptual and theoretical viewpoint on autonomy and how it impacts HRI.

3. Conceptualizing Autonomy

Autonomy has been of philosophical interest for over 300 years. In the 18th century, autonomy was most famously considered by philosopher Immanual Kant as a moral action determined by a person's free will (Kant, 1967). Early psychology behaviorists (e.g., Skinner, 1978) claimed that humans do not act out of free will, rather their behavior is in response to stimuli in the environment. In psychology, autonomy has been primarily discussed in relation to child development, and the term autonomy has been considered as a subjective construct involving self-control, self-governance, and free will. For instance, Piaget (1932) proposed that autonomy is the ability to self-govern as a critical component in a child's moral development. Erikson (1950) similarly defined autonomy as a child's development of a sense of self-control (e.g., early childhood toilet training). In more recent years, attention on autonomy has included the concept of Theory of Mind. Initially centering on child development (Wellman, 1992), this theory now commonly refers to a person's ability to attribute mental states to self and others.

Table 1. Definitions of Autonomy Found in Robotics Literature

Definitions of Agent and Robot Autonomy	
"The robot should be able to carry out its actions and to refine or modify the task and its own behavior according to the current goal and execution context of its task."	Alami et al., 1998, p. 316
"Autonomy refers to systems capable of operating in the real-world environment without any form of external control for extended periods of time."	Bekey, 2005, p. 1
"An autonomous agent is a system situated within and a part of an environment that sense that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future;" "Exercises control over its own actions."	Franklin & Graesser, 1996, p. 25
"An Unmanned System's own ability of sensing, perceiving, analyzing, communicating, planning, decision-making, and acting, to achieve goals as assigned by its human operator(s) through designed HRI … The condition or quality of being self-governing."	Huang, 2004, p. 9
"'Function autonomously' indicates that the robot can operate, self-contained, under all reasonable conditions without requiring recourse to a human operator. Autonomy means that a robot can adapt to change in its environment (the lights get turned off) or itself (a part breaks) and continue to reach a goal."	Murphy, 2000, p. 4
"A rational agent should be autonomous—it should learn what it can to compensate for partial or incorrect prior knowledge."	Russell & Norvig, 2003, p. 37
"Autonomy refers to a robot's ability to accommodate variations in its environment. Different robots exhibit different degrees of autonomy; the degree of autonomy is often measured by relating the degree at which the environment can be varied to the mean time between failures, and other factors indicative of robot performance."	Thrun, 2004, p. 14
"Autonomy: agents operate without the direct intervention of humans or others, and have some kind of control over their actions and internal states."	Wooldridge & Jennings, 1995, p. 116

Autonomy, as a construct representing free will, only encompasses one way in which the term is used. The phenomenon of *psychological* autonomy (and the underlying variables) is different than the phenomenon of *artificial* autonomy that engineers would like to construct in machines and technology (Ziemke, 2008). For instance, when the term autonomy is applied to automation, it is discussed in terms of autonomous function (e.g., performing aspects of a task without human intervention). How is autonomy defined for agents and robots? Robot autonomy has been discussed in the literature as a psychological construct and as an engineering construct. In fact, the term is used to describe many different aspects of robotics, from the robot's ability to self-govern to the level of necessary human intervention. Some definitions of robot autonomy are presented in Table 1.

To clarify the term, we propose the following*:* A generally accepted definition of autonomy is as follows: the extent to which a system can carry out its own processes and operations without external control. This general definition of autonomy can be used to denote autonomous capabilities of humans or machines. However, a stronger and more specific definition can be given to robots by integrating the definitions provided in Table 1. In this paper, we define autonomy as follows:

The extent to which a robot can **sense** its environment, **plan** based on that environment, and **act** upon that environment with the intent of reaching some **task-specific goal** (either given to or created by the robot) without external **control**.

The proposed stronger definition of autonomy integrates current definitions of autonomy and highlights prevalent characteristics of autonomy (i.e., sense, plan, act [see Rosen & Nilsson, 1966 for Hierarchical Paradigm, and Murphy, 2000 for an overview and the Hybrid-Reactive Paradigm], and task-specific goals and control). This definition implies that all of these characteristics of autonomy are important to consider; however, the weighting of each might vary between applications. Note that both the generally accepted and stronger definition begin with the phrase, "the extent to which…" This word choice exemplifies that autonomy is not all or nothing. Autonomy exists on a continuum ranging from no autonomy to full autonomy. Finally, notice the addition of the terminology "task-specific." As discussed later in this framework, defining the robot's autonomy level cannot be considered outside of task-specific context.

3.1 Autonomy in Automation

We first reviewed the human-automation literature to guide our framework of autonomy in HRI. Human-automation researchers have a history of studying and understanding human-related variables, which can be informative for the HRI community. Automation is most often defined as "device or systems that accomplishes (partially or fully) a function that was previously, or conceivably could be, carried out (partially or fully) by a human operator" (Parasuraman, Sheridan, & Wickens, 2000, p. 287). Various taxonomies, classification systems, and models related to levels of automation (LOA) have been proposed. The earliest categorization scheme, which organized automation along both degree and scale, was proposed by Sheridan and Verplank (1978). This 10-point scale categorized higher LOA as representing increased autonomy and lower levels as decreased autonomy (see Table 2). This taxonomy specified what information is communicated to the human (feedback) as well as allocation of function split between the human and automation. However, the scale used in this early taxonomy was limited to a specified a set of discernible points along the continuum of automation applied primarily to the *output* functions of decision-making and action selection. It lacked detailed specification of *input* functions related to information acquisition (i.e., sensing).

Endsley and Kaber (1999) proposed a revised taxonomy with greater specificity on *input* functions, such as how the automation acquires information and formulates options (see Table 3). The Ensley and Kaber model was used to describe each of the automation levels. The taxonomy was organized according to four generic functions, which include: (1) *monitoring*—scanning displays; (2) *generating*—formulating options or strategies to meet goals; (3) *selecting*—deciding upon an option or strategy; and (4) *implementing*—acting out a chosen option.

Table 2. Levels of Decision Making Automation (Sheridan & Verplank, 1978)

Level of Automation	Description
1.	The computer offers no assistance; the human must make all decisions and actions
2.	The computer offers no assistance; the human must make all decisions and actions
3.	The computer offers a complete set of decision/action alternatives, or
4.	Narrows the selection down to a few, or
5.	Suggests one alternative
6.	Executes that suggestion if the human operator approves, or
7.	Allows the human a restricted time to veto before automatic execution, or
8.	Executes automatically, then necessarily informs the human, and
9.	Informs the human only if asked, or
10.	Informs the human only if it, the computer, decides to

Table 3. Levels of Automation (Endsley & Kaber, 1999)

Level of Automation	Description
Manual Control:	The human monitors, generates options, selects options (makes decisions), and physically carries out options.
Action Support:	The automation assists the human with execution of selected action. The human does perform some control actions.
Batch Processing:	The human generates and selects options; then they are turned over to automation to be carried out (e.g., cruise control in automobiles).
Shared Control:	Both the human and the automation generate possible decision options. The human has control of selecting which options to implement; however, carrying out the options is a shared task.
Decision Support:	The automation generates decision options that the human can select. Once an option is selected, the automation implements it.
Blended Decision Making:	The automation generates an option, selects it, and executes it if the human consents. The human may approve of the option selected by the automation, select another, or generate another option.
Rigid System:	The automation provides a set of options and the human has to select one of them. Once selected, the automation carries out the function.
Automated Decision Making:	The automation selects and carries out an option. The human can have input in the alternatives generated by the automation.
Supervisory Control:	The automation generates options, selects, and carries out a desired option. The human monitors the system and intervenes if needed (in which case the level of automation becomes Decision Support).
Full Automation:	The system carries out all actions.

Around the same time Endsley and Kaber proposed their taxonomy, Parasuraman, Sheridan, and Wickens (2000) proposed their conceptual model for types and LOA. Similar to Endsley and Kaber (1999), Parasuraman et al. suggested that functions can be automated to differing degrees along a continuum of low to high (i.e., fully manual to fully automated), and stages of automation represent input and output functions. The stages included: (1) *information acquisition*; (2) *information analysis*; (3) *decision and action selection*; and (4) *action implementation* (Figure 1).

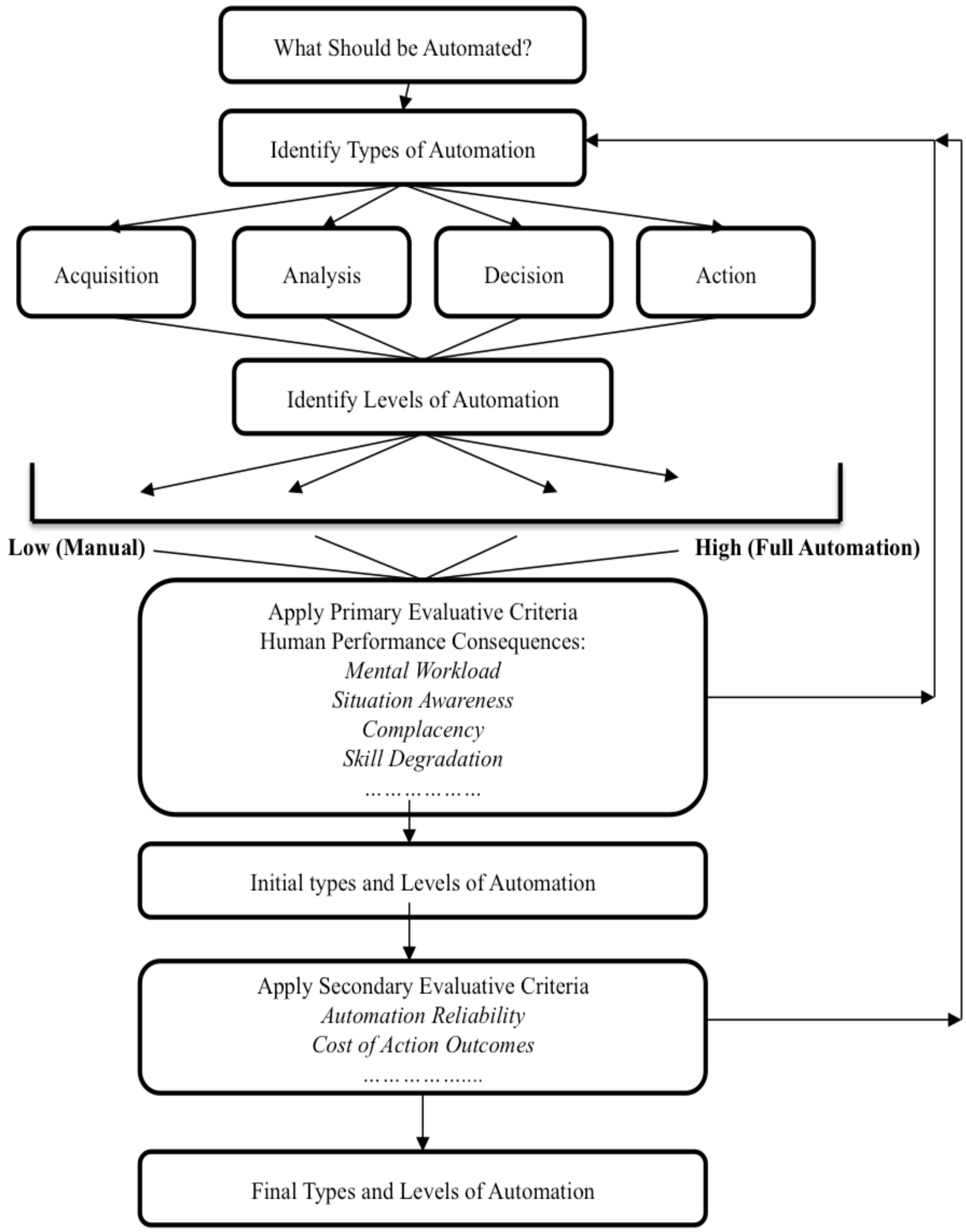

Figure 1. Flow chart showing application of the model of types and levels of automation (adapted with permission from Parasurman, Sheridan, & Wickens, 2000).

Automation categorized under the *information acquisition* stage supports processes related to sensing and registering input data. This stage of automation supports human sensory and perceptual processes, such as assisting humans with monitoring environmental factors. Automation in this stage may include systems that scan and observe the environment (e.g., radar, infrared goggles). At higher levels of information acquisition, automation may organize sensory information (e.g., an automated air traffic control system that prioritizes aircraft for handling). The *information analysis* stage refers to automation that performs tasks similar to human cognitive function, such as working memory. Automation in this stage may provide predictions, integration of multiple input values, or summarization of data to the user. Automation in the *information analysis* stage is different from automation in the *information acquisition* phase, in that the information is manipulated and assessed in some way.

Automation included in the *decision selection* stage selects from decision alternatives. For example, automation in this stage may provide navigational routes for aircraft to avoid inclement weather or recommend diagnoses for medical doctors. Finally, *action implementation* automation refers to automation that executes the chosen action. In this stage, automation may complete all, or subparts, of a task. For example, action automation may include an automatic stapler in a photocopy machine or an autopilot function in an aircraft.

Unlike Endsley and Kaber (1999), Parasuraman, Sheridan, and Wickens (2000) identified primary and secondary evaluative criteria, as depicted at the bottom of the model (Fig. 1). In other words, the purpose of Parasuraman and colleagues' model was to provide an objective basis for making the choice on to what extent a task should be automated. The authors proposed an evaluation of the consequences of both the human operator and the automation. For example, consider decision aid automation at a low level of automation. This automation would be evaluated via the primary evaluative criteria (e.g., human performance, such as workload, situation awareness), and then the level of automation is adjusted (e.g., higher LOA of the decision aid automation would reduce the workload). Next, secondary criteria are evaluated (e.g., automation reliability, cost of decision/action outcomes), and again, the level of automation is adjusted (e.g., if the LOA is above a certain level, then reliability decreases). This iterative method provides a process for determining appropriate levels or ranges of automation to identify potential design issues.

3.2 Benefits and Limitations of Applying Automation Autonomy Models to Human-Robot Interaction

Next we consider how autonomy has been conceptualized in HRI and why a new model is needed. Why not directly apply the automation models and taxonomies (Endsley & Kaber, 1999; Parasuraman, Sheridan, & Wickens, 2000; Sheridan & Verplank, 1978) to robotics? The answer is that these models can inform HRI but only to a certain point. Each model provides an organizational framework in which to categorize not only the purpose or function of the automation (e.g., stages), but also considers automation along a continuum of autonomy. These models are important to consider within the context of both robotics and HRI, because they can serve as a springboard for development of similar taxonomies and models specific to robot autonomy. In particular, Sheridan and Verplank's taxonomy (Table 2) has been suggested as appropriate to potentially describe a robot's level of autonomy (Goodrich & Schultz, 2007).

However, considering differences between automation and robotics is important. Capabilities such as mobility, environmental manipulation, and social interaction separate robots from automation in both function and physical form. The goal here is not to redefine robot or automation, rather simply to depict that robots are a technology class of their own, separate but related to automation. Robots may serve different functions relative to traditional automation; for example, some (but certainly not all) robots may play a social role. Social ability is not a construct considered in the LOA models and taxonomies. A complementary way to think about how these taxonomies could relate to HRI is to consider the degree to which the human and robot interact and to what extent each can act autonomously. The next sections address how autonomy has been applied to HRI and how autonomy's conceptualization in HRI is similar or different from human-automation interaction.

3.3 Autonomy in Human-Robot Interaction

Autonomy within an HRI context is a widely considered construct; however, ideas surrounding how autonomy influences human-robot interaction are varied. There are two schools of thought with which autonomy and HRI have been conceptualized: (1) higher robot autonomy requires *less frequent HRI*; and (2) higher robot autonomy requires *higher levels of HRI*.

The first viewpoint, that higher autonomy requires less HRI, has been supported by Huang and colleagues (Huang et al., 2004; Huang, Pavek, Albus, & Messina, 2005; Huang, Pavek, Novak, Albus & Messina, 2005; Huang et al., 2007). Their goal was to develop a framework for autonomy and metrics used to measure robot autonomy. Although this framework is used primarily within military applications, the general framework has been cited more generally as a basis for HRI autonomy (Yanco & Drury, 2004).

In the Huang framework, the relationship between the level of HRI and the autonomy level of the robot "…is fairly linear for simple systems" (Huang et al., 2004, p. 5). They proposed a negative linear correlation between autonomy and frequency of HRI, so that as the level of robot autonomy (LORA) increases, the HRI frequency decreases (see Fig. 2). Their model included constructs such as human intervention (number of unplanned interactions), operator workload (as measured by NASA TLX), operator skill level, and the operator-to-robot ratio.

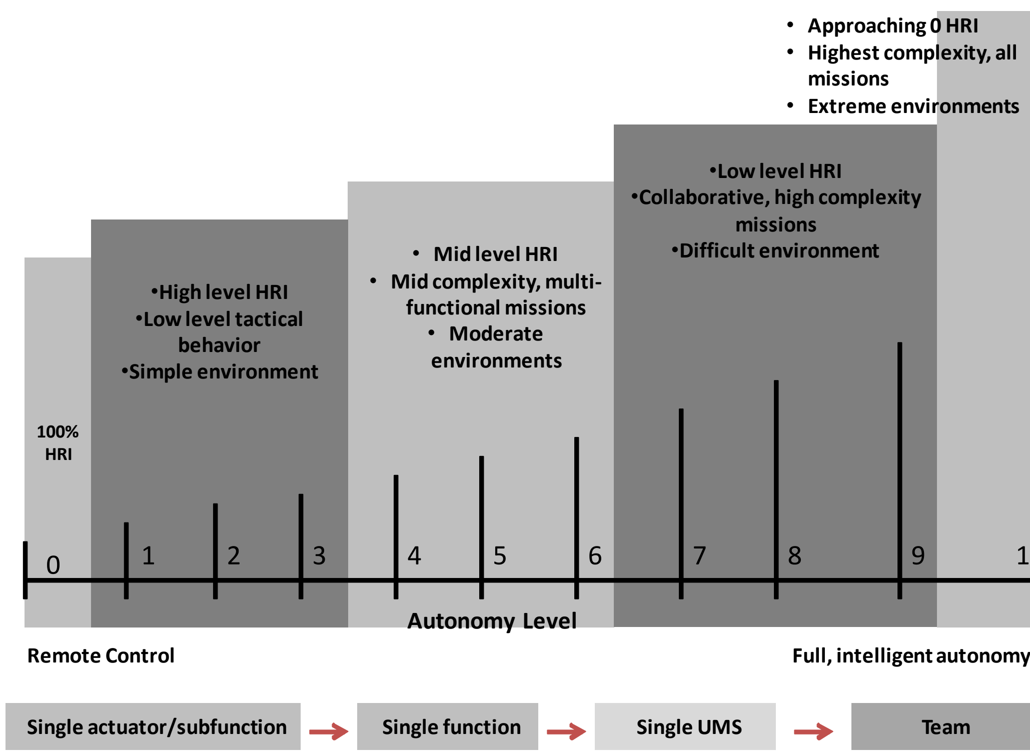

Figure 2. Autonomy Levels for Unmanned Systems (ALFUS) model of autonomy, depicting level of HRI along the autonomy continuum (adapted with permission from Huang, Pavek, Albus, & Messina, 2005). Note: Unmanned System (UMS).

Other researchers have also proposed that higher robot autonomy requires less interaction (Yanco & Drury, 2004). Autonomy has been described as the amount of time that a person can neglect the robot, and neglect time refers to the measure of how the robot's task effectiveness

(performance) declines over time when the robot is neglected by the user (Goodrich & Olsen, 2003). Robots with higher autonomy levels can be neglected for a longer time periods.
“There is a continuum of robot control ranging from teleoperation to full autonomy: the level of human-robot interaction measured by the amount of intervention required varies along this continuum. Constant interaction is required at the teleoperation level, where a person is remotely controlling the robot. *Less interaction is required as the robot has greater autonomy*” [emphasis added] (Yanco & Drury, 2004, p. 2845).

The idea that higher autonomy reduces the frequency of interaction is a stark contrast to the other school of thought in which HRI researchers have proposed that more robot autonomy enables more sophisticated interaction (e.g., Thrun, 2004; Feil-Seifer, Skinner, & Mataric, 2007; Goodrich & Schultz, 2007). Thrun’s (2004) framework of HRI defined categories of robots, and each category required a different level of autonomy as dictated by the robot’s operational environment. Professional service robots (e.g., museum tour guides, search and rescue robots) and personal service robots (e.g., robotic walkers) mandate higher degrees of autonomy because they operate in a variable environment and interact in close proximity to people. Thrun declared, "human-robot interaction cannot be studied without consideration of a robot's degree of autonomy, because it is a determining factor with regards to the tasks a robot can perform and the *level* at which the interaction takes place" [emphasis added] (2004, p. 14).

Furthermore, autonomy has been proposed as a benchmark for developing *social interaction* in socially assistive robotics (Feil-Seifer, Skinner, & Mataric, 2007). Here, autonomy serves two functions: (1) to perform well in a desired task and (2) to be proactively social. However, Feil-Seifer et al. warned that the robot’s autonomy should allow for social interaction only when appropriate (i.e., only when social interaction enhances performance). Developing autonomous robots that engage in peer-to-peer collaboration with humans may be harder to achieve than high levels of autonomy with no social interaction (e.g., iRobot Roomba; Goodrich & Schultz, 2007).

Indeed, an important distinction between the two conceptualizations of autonomy is the role of the robot and the human in the interaction (Johnson, et al., 2010; Murphy & Schreckenghost, 2013; Scholtz, 2002a; Scholtz, 2002b). Consider the use of two terms that define the human’s role in relation to robot autonomy: intervention and interaction. Conceivably, intervention could be interpreted as a specific type of interaction (as suggested in Huang et al., 2004), referring to the frequency of human control. However, having a robot act autonomously with no intervention mirrors the human out of the loop phenomenon in automation, which is known to cause performance problems (e.g., Endsley, 2006; Endsley & Kiris, 1995). One might also consider the frequency of intervention more applicable when the human’s role is to operate the robot (e.g., teleoperation or monitoring). On the other hand, the sophistication of the interaction might be more applicable when the human’s role is that of a bystander or peer (e.g., social partner, coworker, or supervisor). In sum, it is important to simultaneously consider both intervention and interaction when determining a robot’s level of autonomy by asking how much and what level of interaction is required.

In summary, a framework of autonomy in HRI is needed. As this literature review revealed, autonomy is an important construct related to automation and HRI, albeit the term has been conceptualized differently between and within these two fields. We moved toward developing the building blocks for a framework of robot autonomy—a framework that is influenced by the automation models (Endsley & Kaber, 1999; Parasuraman, Sheridan, & Wickens, 2000; Sheridan & Verplank, 1978) while also taking into consideration the unique aspects of robot autonomy (Feil-Seifer, Skinner, & Mataric, 2007; Goodrich & Olsen, 2003; Goodrich & Schultz, 2007; Huang, Pavek, Albus, & Messina, 2005; Thrun, 2004; Yanco & Drury, 2004) that may be different from automation.

The integration of previous work will move the field toward a more cohesive definition and conceptualization of autonomy. In the next sections, the framework guidelines are proposed to inform designers and researchers to consider what level of autonomy is appropriate for their robot and the impact of autonomy on HRI. HRI-specific guidelines were lacking in previous work, and we will move toward framing autonomy in a way that is more easily understood, more easily applied to defining autonomy of robots, and HRI-specific.

4. Toward a Framework of Levels of Robot Autonomy (LORA) in Human-Robot Interaction

We provide a framework for examining LORA in relation to HRI. This framework is highly influenced by the taxonomies and models of LOA (Endsley & Kaber, 1999; Parasuraman, Sheridan, & Wickens, 2000). However, we have proposed several key changes and additions to the way in which autonomy should be conceptualized for HRI. First, we highlight the importance of autonomy to be determined within the context of task and environment. This is highlighted because robots are situated in, and typically physically manipulate, their environment. Second, we propose a taxonomy for categorizing robot autonomy, which can be seen as lying along a continuum. This taxonomy is, in part, adapted from Endsley and Kaber (1999) with differences in terminology and some definitions. These differences are important because this taxonomy is meant to be HRI-specific (e.g., "teleoperation" instead of "action support"). Third, different from the Parasuraman et al. (2000) model, we highlight specific HRI variables that are likely to be influenced by robot autonomy. Some variables differ from automation frameworks, such as the addition of social constructs. Fourth, we rethink the notion of function allocation (as rigidly depicted in Sheridan and Verplank, 1978), recognizing that sliding, fluid, or changing autonomy is likely to occur with semi-autonomous robots. We stress that a robot's autonomy may fluctuate; thus the autonomy level, as categorized in the taxonomy, may change depending on the environment, task, and interaction over time. This point is highlighted in the thought experiments presented at the end of the framework.

This framework includes guidelines that a designer or researcher may use when considering autonomy in HRI (Figure 3). Guidelines 1-3 serve to determine robot autonomy. Guideline 4 categorizes robot autonomy via a taxonomy. Finally, Guideline 5 broadly suggests the implications of the robot autonomy on HRI (i.e., human variables, robot variables, and interaction variables). The next sections describe each of these guidelines in more detail.

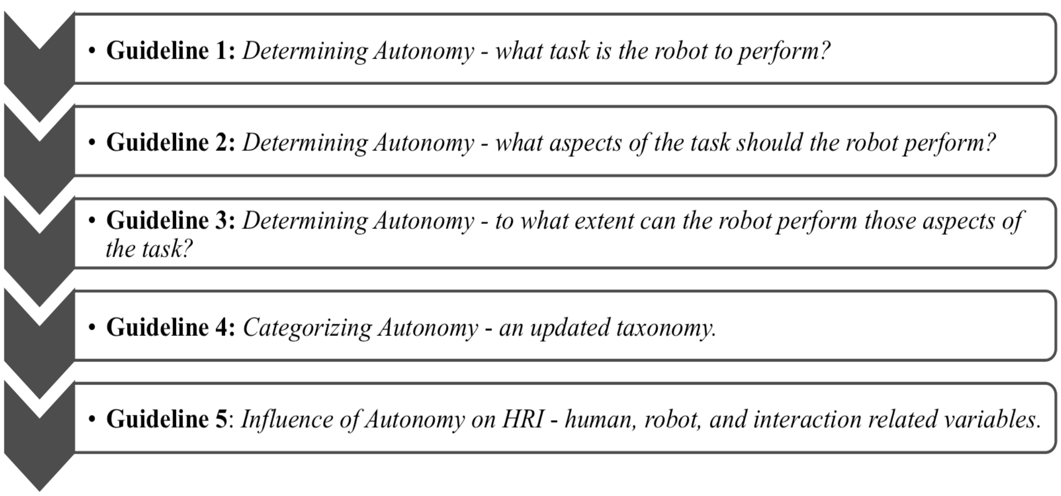

Figure 3. Organizational flow chart to determine robot autonomy and effects on HRI.

4.1 Determining Robot Autonomy (Guidelines 1–3)

Guidelines 1–3 are meant to provide suggested steps for determining and measuring robot autonomy. Specifically, the proposed guidelines in this section focus on HRI, with an emphasis on function allocation between a robot and a human. Consideration of the task and environment is particularly important for robotics, because robots are embodied, that is, they are situated within an environment and usually expected to perform tasks by physically manipulating that environment. A robot's capability to sense, plan, and act within its environment is what determines its autonomy. Therefore, in this framework, the first determining question to ask is, "What task is the robot to perform?"

The robot designer should not ask, "Is this robot autonomous?"; rather, the important consideration is, "Can this robot complete the given task at some level of autonomy?" For instance, the iRobot Roomba is capable of navigating and vacuuming floors autonomously. However, if the task of vacuuming is broadened to consider other subtasks (i.e., picking up objects from the floor, cleaning filters, emptying dirt bins or bags), then the Roomba may be considered semi-autonomous because it only completes a portion of those subtasks. Likewise, if the environment is changed (e.g., vacuuming stairs opposed to flat surfaces), then the Roomba's autonomy could be categorized differently, since it is currently incapable of vacuuming stairs. Therefore, specifying the context of the task and environment is critical for determining the task-specific LORA.

Secondly, specifying demands, such as task criticality, accountability, organizational structure, and environmental complexity, should guide a designer in demining autonomy. Task characteristics and consequences of error have been shown to be influenced by automation level (Carlson, Murphy, & Nelson, 2004). In many cases, failures or errors at early stages of automation are not as critical as errors at later stages of automation. One rationale is that it may be risky to program a machine to have high autonomy in a task that requires decision support, particularly if the decision outcome involves lethality or other human safety concerns (Parasuraman, Sheridan, & Wickens, 2000; Parasuraman & Wickens, 2008). For example, unreliability in a robot that autonomously navigates in an office environment may result in either false alarms or misses of obstacles. In this example, the criticality of errors is substantially less than errors conducted by a robot that autonomously navigates in a search and rescue task (Casper & Murphy, 2003; Lewis et al., 2010). Here, navigational errors could be detrimental due to risk of causing secondary collapses in unstable structures. Another example is a robot that determines what medication a patient should take. In these two examples, robot failure may result in critical, if not lethal, consequences.

Accountability for successful task completion is of consideration, particularly as robots and humans work as teams. As robots become more autonomous and are perceived as peers or teammates, it is possible that the distribution of responsibility may be perceived to be split between the robot and the human. Robot autonomy has been shown to play a role in participants' accountability of tasks errors. When a robot is perceived as more autonomous, participants reported less self-blame (accountability) for task errors (Kim & Hinds, 2006); thus, responsibility of consequences may be misplaced and the human operator may feel less accountable for errors. In fact, healthcare professionals have reported concern for who (the professional or a medical robot) may be accountable for medical errors (Tiwari, Warren, Day, & MacDonald, 2009). Therefore, care should be taken in determining which tasks a robot will perform, as well as in designing the system so that human supervisors are held accountable and can easily diagnose and alleviate consequences of error.

Furthermore, the nature of the environment should be considered. In an ethnographic study, an autonomous robot was placed in a workplace environment; how the robot was used impacted the social workflow and other organizational factors tied to the environment (Mutlu & Forlizzi, 2008). Service robots designed for assistive functions (e.g., home, workplace, or healthcare applications), surveillance, or first responders (e.g., search and rescue) will be required to operate in unknown, unstructured, and dynamic environments, which will certainly influence the functional requirements of the robot. The robot's capability to operate in a dynamic environment is highly dependent on environmental factors (e.g., lighting, reflectivity of surfaces, glare) that influence the robot's sensors to perceive the world around it. Higher LORA may be required for a service robot to function in dynamic, ever-changing environments (Thrun, 2004). However, not all aspects of the environment can be anticipated; thus, for many complex tasks, the presence of a human supervisor may be required (Desai, Stubbs, Steinfeld, & Yanco, 2009).

Once the task and environmental demands are determined, the next question is, “What aspects of the task should the robot perform?” Each task, no matter how simple or complex, can be divided into primitives: sense, plan, and act (Murphy, 2000; Rosen & Nilsson, 1966). Consider robots equipped with assisted teleoperation features (e.g., Takayama et al., 2011): A teleoperated robot demonstrates low levels of autonomy by assisting the human operator in obstacle avoidance. Usually, this feature is programmed into the robot architecture using behavior-based sense-act couplings (e.g., behavior-based robotics; Arkin, 1998), where the robot is assisting with the aspects of the task by detecting obstacles (sense), then adjusting its behavior to avoid collision (act). The human remains, in large part, in charge of path planning and navigational goals (plan). However, a robot that navigates semi-autonomously (e.g., Few et al., 2008) may require a human to specify the high-level goal of navigating to a specified location. Once the high-level goal is given, the robot can then autonomously navigate to that location. Here, the robot demonstrates a high level of autonomy in sensing the environment (sense), a relatively high level of autonomy in the plan primitive (except for the human provided the high-level goal), and a high level of autonomy in physically implementing objectives toward the goal (act).

As these examples suggest, autonomy can vary along any of the sense, plan, and act primitives (Murphy, 2000; Rosen & Nilsson, 1966), which relates to the next determining question, “To what extent can the robot perform these aspects of the task?” Each of the sense, plan, and act primitives could be allocated to either the human or the robot (or both). Similar to the Parasuraman, Sheridan, and Wickens’ (2000) stages of automation, a robot can vary in autonomy level (from low to high) along the three primitives (see Figure 4).

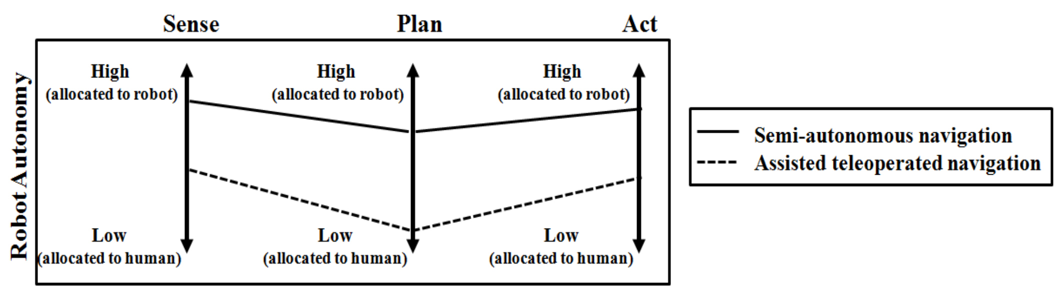

Figure 4. Levels of autonomy across the robot primitives sense, plan, and act. Two examples are given: assisted teleoperation (dotted line) and semi-autonomous navigation (solid line). Model modified from Parasuraman, Sheridan, and Wickens, 2000.

As depicted in Figure 4, the level of autonomy may vary from low to high for each of the robot primitives. Allocation along the three primitives does not imply that autonomy should be thought of as an additive model that would collapse these different amounts of autonomy into one scale. Instead, this depiction (Fig. 4) represents the fact that a robot might vary in capability along different aspects of the task. Determining the robot’s autonomy prompts the need for a descriptive clarification of how to measure the extent or degree to which a robot can perform each task aspect (sense, plan, or act). Levels of autonomy are most often identified by function allocation in the automation literature. For example, in the Endsley and Kaber (1999) model, the level of automation is specified through a taxonomy that is based on the allocation of function to either the human or the automation.

In HRI, function allocation has been commonly measured by amount of human intervention (Yanco & Drury, 2004). Specifically, human intervention is measured by the percentage of time a task is completed on its own, and intervention is measured by the percentage of time the human must control the robot. The two measures, autonomy and intervention, must sum to 100 percent.

For example, a teleoperated robot has 0 percent autonomy and 100 percent intervention. A fully autonomous robot has 100 percent autonomy and 0 percent intervention. Between these two anchor points lies a continuum of shared control. For example, a medication management robot may select a medication and handoff the medication to a human, but the human might be responsible for high-level directional (navigation) commands. Here, the robot has 75 percent autonomy and 25 percent intervention. Similarly, autonomy has been measured as human neglect time (Olsen & Goodrich, 2003). In this metric, autonomy is measured by the amount of time that the robot makes progress toward a goal before dropping below an effective reliability threshold or requiring user instruction.

Although this idea of measuring the time of intervention and neglect is useful, it has limitations. Amount of time between human interventions may vary as a result of other factors, such as inappropriate levels of trust (i.e., misuse and disuse), social interaction, task complexity, robot capability (e.g., robot speed of movement), usability of the interface or other control method, and response time of the user. Therefore, if interaction time is used as a quantitative measure, care should be taken when generalizing findings to other robot systems or tasks. Furthermore, the notion of function allocation may be difficult to conceptualize for semi-autonomous robots. Semi-autonomous states may include a fusion of human and robot control, where control might be fluid or "sliding" from one autonomy level to another. Thus, in categorizing a robot along a continuum, one should be mindful that the level of autonomy could change—in which case, the robot's autonomy would be better described as a range of levels, rather than as a discrete level.

We propose a supplemental metric be used, such as a qualitative measure of intervention level (i.e., subjective rating of the human intervention) or a general quantitative measure focused on subtask completion, in addition to time (i.e., number of subtasks completed by robot divided by the number of total subtasks required to meet a goal). Each metric has tradeoffs but could provide some general comparative indication of the robot's degree of autonomy.

As we discussed earlier, *intervention* and *interaction* are not necessarily interchangeable terms. Intervention is a type of interaction specific to task-sharing where the human performs some aspect of the task. Interaction may include other factors not necessarily specific to the intervention of task completion, such as verbal communication, gestures, or emotional expression. Some autonomous service robots could work in isolation, requiring little interaction of any kind (e.g., an autonomous pool cleaning robot); whereas other robots working autonomously in a social setting may require a high level of interaction (e.g., an autonomous robot serving drinks at a social event). Finally, the measure of autonomy is specifically applicable to service robots that perform tasks. Neglect time may not be an appropriate measure of autonomy for robots designed for entertainment, for example. Other types or classes of robots may require different evaluative criteria for determining autonomy, which will require extensions of the present framework.

4.2 Categorizing Autonomy: A Taxonomy of Levels of Robot Autonomy for HRI (Guideline 4)

The purpose of the next guideline is to categorize the robot's autonomy using the proposed taxonomy (Table 4). Under-specification of intermediate autonomy levels is a limitation in previous HRI frameworks (e.g., Huang, Pavek, Albus, & Messina, 2005; Yanco & Drury, 2004). For example, if autonomy is measured as a precise level between 0 and 100 percent, what is the difference between 52 percent and 54 percent? Autonomy may be considered along a continuum (e.g., 0–100), but actually conceptualizing specific degrees of autonomy is difficult. In this context, we believe it is valuable to apply conceptual descriptions of categories for different levels of autonomy. In Table 4, we propose a taxonomy influenced, in part, by Ensley and Kaber (1999), in which the robot autonomy can be categorized into conceptual descriptions of "levels of robot autonomy" (LORA).

Table 4. Proposed Taxonomy of Levels of Robot Autonomy for HRI

LORA	Sense	Plan	Act	Description	Examples from Literature
Manual	H	H	H	The human performs all aspects of the task including sensing the environment, generating plans/options/goals, and implementing processes.	"Manual Control" Endsley & Kaber, 1999
Tele-operation	H/R	H	H/R	The robot assists the human with action implementation. However, sensing and planning is allocated to the human. For example, a human may teleoperate a robot, but the human may choose to prompt the robot to assist with some aspects of a task (e.g., gripping objects).	"Action Support" Endsley & Kaber, 1999; Kaber et al., 2000; "Manual Teleoperation" Milgram, 1995; "Tele Mode" Baker & Yanco, 2004; Bruemmer et al., 2005; Desai & Yanco, 2005
Assisted Tele-operation	H/R	H	H/R	The human assists with all aspects of the task. However, the robot senses the environment and chooses to intervene with task. For example, if the user navigates the robot too close to an obstacle, the robot will automatically steer to avoid collision.	"Assisted Teleoperation" Takayama et al., 2011; "Safe Mode" Baker & Yanco, 2004; Bruemmer et al., 2005; Desai & Yanco, 2005
Batch Processing	H/R	H	R	Both the human and robot monitor and sense the environment. The human, however, determines the goals and plans of the task. The robot then implements the task.	"Batch Processing" Endsley & Kaber, 1999; Kaber et al., 2000
Decision Support	H/R	H/R	R	Both the human and robot sense the environment and generate a task plan. However, the human chooses the task plan and commands the robot to implement actions.	"Decision Support" Endlsey & Kaber, 1999; Kaber et al., 2000
Shared Control With Human Initiative	H/R	H/R	R	The robot autonomously senses the environment, develops plans and goals, and implements actions. However, the human monitors the robot's progress and may intervene and influence the robot with new goals and plans if the robot is having difficulty.	"Shared Mode" Baker & Yanco, 2004; Bruemmer et al., 2005; Desai & Yanco, 2005; "Mixed Initiative" Sellner et al., 2006; "Control Sharing" Tarn et al., 1995
Shared Control With Robot Initiative	H/R	H/R	R	The robot performs all aspects of the task (sense, plan, act). If the robot encounters difficulty, it can prompt the human for assistance in setting new goals and plans.	"System-Initiative" Sellner et al., 2006; "Fixed-Subtask Mixed-Initiative" Hearst, 1999
Executive Control	R	H/R	R	The human may give an abstract high-level goal (e.g., navigate in environment to a specified location). The robot autonomously senses environment, sets the plan, and implements action.	"Seamless Autonomy" Few et al., 2008; "Autonomous mode" Baker & Yanco, 2004; Bruemmer et al., 2005; Desai & Yanco, 2005
Supervisory Control	H/R	R	R	The robot performs all aspects of task, but the human continuously monitors the robot, environment, and task. The human has override capability and may set a new goal and plan. In this case, the autonomy would shift to executive control, shared control, or decision support.	"Supervisory Control" Endsley & Kaber, 1999; Kaber et al., 2000
Full Autonomy	R	R	R	The robot performs all aspects of a task autonomously without human intervention with sensing, planning, or implementing action.	"Full Automation" Endsley & Kaber, 1999

*Note: H = Human, R = Robot. Manual represents a situation where no robot is involved in performing the task; this level is included for a complete taxonomy continuum.

The taxonomy has a basis in HRI by specifying each LORA from the perspective of the interaction between the human and robot and the roles that they each play. That is, for each proposed LORA, we (1) suggest function allocation between robot and human for each of the sense, plan, and act primitives (see Fig. 4; Murphy, 2000; Rosen & Nilsson, 1966); (2) provide a proposed description of each LORA; and (3) illustrate with examples of service robots from the HRI literature that provide an approximate representation of the categorical description. Table 4 includes a mix of empirical studies involving robots and simulations, as well as published robot autonomy architectures.

It is important to note that autonomy is a continuum, thus there are blurred borders between the proposed categories. Moreover, the taxonomy represents a range of feasible categories of robot autonomy. There may be other combinations of sense, plan, and act that are not included in this taxonomy. These alternate combinations are certainly not ruled out but rather excluded here, because they are technically or practically uncommon or unlikely. Furthermore, there are some robots that perform a subset of these primitives (e.g., sense and act only). Thus, levels should not be treated as exact descriptors of a robot's autonomy but rather as an approximation of a robot's autonomy level along the continuum. The taxonomy is not meant to be a complete or exhaustive set of categories, rather we are proposing an example set of categories which could be seen as lying along the autonomy continuum. They are qualitative, not quantitative, because they are descriptive; therefore, the taxonomy levels are not necessarily representative of a scale but rather suggestive of a possible ordering of categories.

Lastly, we recognize that due to the complex nature of HRI, it is likely a robot's autonomy level may fluctuate or change throughout an interaction within any given environment, task, or interaction. Thus, to consider sliding autonomy, there is a constant need to reconsider autonomy level categorization throughout duration of use.

4.3 The Influence of Autonomy on HRI (Guideline 5)

The last guideline of the framework was to evaluate the influence of the robot's autonomy on HRI. Research on automation and HRI provides insights for identifying variables influenced by robot autonomy. The framework includes variables related to the human and to the robot (Beer, Fisk, and Rogers, 2012). This listing is not exhaustive. Many other variables (e.g., safety, control methods, robot appearance, perceived usefulness) might also influence, and be influenced by, robot autonomy and in need of further investigation. Our proposed framework provides evaluation criteria to examine the interaction between autonomy and HRI-related variables and determine if the robot's autonomy level is appropriate for supporting optimal human-robot interaction.

4.3.1 Robot-related variables

The robot's intelligence and learning capabilities are important to consider along the autonomy continuum because both of these variables influence what and how the robot performs. Not all robots are intelligent, but robots that demonstrate higher levels of autonomy for complex tasks may require higher intelligence. According to Bekey (2005), robot intelligence may manifest as sensor processing, reflex behavior, special purpose programming, or cognitive functioning. Generally speaking, the more autonomous a robot is, the more sophisticated these components may be. In the future, it is expected that most autonomous robots will be equipped with some ability to learn. This will be especially true as robots are moved from the laboratory to an operational environment, where the robot will have to react and adjust to unpredictable and dynamic environmental factors (Bekey, 2005; Russell & Norvig, 2003; Thrun, 2003).

As robots move from the laboratory to more dynamic environments (e.g., the home, hospital setting, workplace), reliability is generally expected to be less than perfect because of constraints in designing algorithms to account for every possible scenario (Parasuraman & Riley, 1997). Reliability should be measured against a threshold of acceptable error but how best to determine the appropriate threshold? Addressing this question proves to be a balancing act between designing with the assumption that the machine will sometimes fail and consideration for how such failures will affect human performance. In automation, degraded reliability at higher levels of autonomy resulted in an “out of the loop” operator (Endsley, 2006; Olsen & Goodrich, 2003), where the operator may be unable to diagnose the problem and intervene in a timely manner (i.e., extended time to recovery; Endsley & Kaber, 1999; Endsley & Kiris, 1995). To reduce “out of the loop” issues and contribute to the user’s recognition of the robot’s autonomy level, developers should design the robot to allow the user to understand what the robot is doing. Automated tasks where an operator can form a mental model are referred to as transparent. Increased autonomy was reported as a problem due to lack of transparency in a remote rover field study (Stubbs, Hinds, & Wettergreen, 2007). A robot that provides adequate feedback about its operation may achieve transparency. However, much consideration is needed in determining how much, when, and what type of feedback is most beneficial for a given task and robot autonomy level.

4.3.2 Human-related variables

Situation awareness (SA) and mental workload have a long history in the automation literature. These concepts are inherently intertwined (see, Tsang & Vidulich, 2006) and empirical evidence suggests that both influence human performance changes as a function of LOA. SA is “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” (Endsley, 1995, p. 36). Mental workload is “the relation between the function relating the mental resources demanded by a task and those resources available to be supplied by the human operator” (Parasuraman, Sheridan, & Wickens, 2008, pp. 145-146). An imbalance between SA and workload can lead to performance errors. The relationship between workload, SA, and LOA is complex but generally negative: as LOA increases workload decreases and vice versa. However, low workload during high LOA may lead to boredom (Endsley & Kiris, 1995), particularly in monitoring tasks (e.g., air traffic control). On the other end of the spectrum, high workload during low LOA generally leads to low operator SA and decreased performance (Endsley & Kaber, 1999; Endsley & Kiris, 1995).

The rich empirical background of SA and workload in the automation literature can inform robotics. Although the automated systems evaluated have been primarily studied in the context of air traffic control and aviation, similar human-machine interactions may be expected in HRI. In fact, much of the work involving SA and robotics has been conducted in similarly dynamic service environments and tasks (e.g., Kaber, Onal, & Endsley, 2000; Kaber, Wright, & Sheik-Nainar, 2006; Riley & Endsley, 2004; Scholtz, Antonishek, & Young, 2004; Sellner, Heger, Hiatt, Simmons, & Singh, 2006). SA at low levels of autonomy may primarily focus on where the robot is located, what obstacles lay ahead, or deciphering the sensor data the robot produces. As a robot approaches higher autonomy levels, it may be perceived as a teammate or peer (Goodrich & Schultz, 2007; Milgram, Rastogi, & Grodski, 1995). SA associated with a robot peer may more closely resemble that of shared SA, where the human must know the robot’s status, and likewise, the robot must know the human’s status to the degree that they impact each other’s tasks and goals. Design principles for supporting SA in team operations (Endsley, Bolte, & Jones, 2003; Gorman, Cook, & Winner, 2006) may be applied to human-robot teams and need to be empirically tested.

Other human-related variables, such as trust and acceptance, have been increasingly studied within the context of HRI (Desai, Kaniarasu, Medvedev, Steinfeld, & Yanco, 2013; Desai et al., 2012). A number of models and theories related to trust in automation (Cohen, Parasuraman, & Freeman, 1998; Dzindolet et al., 2003; Lee & See, 2004; Madhavan & Wiegmann, 2007) and preliminary frameworks of trust in HRI have been proposed (Desai, Stubbs, Steinfeld, & Yanco, 2009; Hancock, Billings, & Schaefer, 2011). These models suggest that trust, in conjunction with many other factors, can predict robot use. It is important to consider how the nature of trust may vary along the autonomy continuum. For example, trust in a teleoperated system (e.g., the sensors are reliable, the feedback is accurate, the robot will respond to controls) might be very different from trust in a more autonomous system (e.g., when the robot is fulfills a teammate role).

Although the frameworks of trust in HRI have borrowed from the automation literature, there are some important differences to consider that are in need of empirical evaluation. First, automation generally lacks physical embodiment (i.e., many automated systems are primarily software-based). Many robots are physically mobile, look or behave like humans or animals, and physically interact with the world. Physical robot characteristics (e.g., size, weight, speed of motion) and their effects on trust need to be empirically evaluated. Second, unlike most automated systems, some service robots are designed to be perceived as teammates or peers with social capabilities, rather than as tools (e.g., Breazeal, 2005; Groom & Nass, 2007). Understanding how to develop trust in robots is an avenue of research critical for designing robots meant to be perceived as social partners.

As robots become increasingly advanced and perform complex tasks, the robot's autonomy will be required to adjust or adapt between levels. In general, robotic and automated systems that operate under changing levels of autonomy (e.g., switching between intermediate levels) are not addressed in the trust literature. Many avenues of research need to be pursued to better understand the role of trust in HRI, how trust in robots is developed, and how misuse and disuse of robots can be mitigated.

Acceptance is also an important human-related variable to consider with regard to predicting technology use (Davis, 1989), as well as HRI outcomes (Broadbent, Stafford, & MacDonald, 2009; Young, Hawkins, Sharlin, & Igarashi, 2009). Designers should be mindful that radical technologies such as personal robots are not as readily accepted as incremental innovations (Dewar & Dutton, 1996; Green, Gavin, & Aiman-Smith, 1995). Despite the research community's acknowledgement that acceptance is an important construct, further investigation is needed to understand and model variables that influence robot acceptance, how such variables interact, and finally, how predictive value varies over the autonomy continuum.

4.3.3 Social-related variables

Not all service robots are social. However, for those robots that are, designers should consider how social interaction relate to autonomy. Robots are one of the few technologies in which design has been modeled in part by science-fiction portrayals of autonomous systems (Brooks, 2002). Even though most individuals of the general population have never interacted with a robot directly, most people have ideas or definitions of what a robot should be like (Ezer, Fisk, & Rogers, 2009; Khan, 1998). If users have preconceived notions of how robots should socially behave, then it becomes all the more important to understand how to match user expectations with the robot's actual autonomy. According to Breazeal (2003), when designing robots, the emphasis should not be on whether people will develop a social model to understand robots. Instead, it is more important that the robot adhere to the social models that humans expect. What social models do people hold for robots? And do those social models change as a function of robot autonomy?

The research community generally accepts that people treat technologies as social actors (Nass, Fogg, & Moon, 1996; Nass & Moon, 2000; Nass, Moon, Fogg, Reeves, 1995; Nass, Steuer, Henriksen, & Dryer, 1994), particularly occurring with humanoid robots (Breazeal, 2005). Social capability has been categorized into classes of social robots (Breazeal, 2003; Fong, Nourbakhsh, & Dautenhahn, 2003): socially evocative, social interface, socially receptive, sociable, socially situated, socially embedded, and socially intelligent. These classes can be considered as a continuum (from socially evocative, where the robot relies on human tendency to anthropomorphize, to socially intelligent, where the robot shows aspects of human-style social intelligence, based on models of human cognition and social competence). Social classes higher on this continuum require greater amounts of autonomy to support the complexity and effectiveness of the HRI.

It is difficult to determine the most appropriate metric for measuring social effectiveness. A variety of metrics have been proposed (Steinfeld et al., 2006) and applied via interaction characteristics (e.g., interaction style or social context), persuasiveness (i.e., robot is used to change the behavior, feelings, or attitudes of humans), trust, engagement (sometimes measured as duration), and compliance. Appropriate measures of social effectiveness may vary along the autonomy continuum. For instance, when a robot is teleoperated, definitive social interaction may not exist between the robot and human. In fact, the robot may be designed to facilitate social communication between people (i.e., the operator and a remotely located individual; Lee & Takayama, 2011; Tsui, Desai, Yanco, & Uhlik, 2011). In this case, "successful social interaction" may be assessed by the quality of remote presence (the feeling of the operator actually being present in the robot's remote location). Proper measures of "social effectiveness" may be dictated by the quality of the system's video and audio input/output, as well as communication capabilities, such as lag time or delay, jitter, or bandwidth (Steinfeld et al., 2006). Social interaction with intermediate or fully autonomous robots may be more appropriately assessed by the social characteristics of the robot itself (e.g., appearance, emotion, personality; Breazeal, 2003; Steinfeld et al., 2006).

5. A Framework of Robot Autonomy

A graphical representation of the framework, specifically the taxonomy and guidelines, is depicted in Figure 5. From top to bottom, the figure depicts the five guidelines. This framework is not meant to present a method. Rather, the framework should be treated as a suggested guide to determine autonomy, categorize the LORA using a taxonomy, and consider which HRI variables may be influenced by the LORA. These guidelines might be considered in the order presented; however, there are likely instances where the guidelines should be considered in a different order, or certain guidelines need to be reconsidered at different stages of the design cycle.

To illustrate how this framework could be applied, we present two thought experiments. First, consider the design of service robots to assist with medication management for older adult users. The first consideration may be the task and environmental variables. In this example, task criticality (e.g., high—the correct medication must be chosen), environment complexity (e.g., medium—homes vary), and task accountability (e.g., uncertainty whether mistakes would be attributed to the robot or to the user) are all relevant variables. Next, subcomponents of the task may be considered. For example, a particular robot may reliably deliver medication but less reliably select a particular medication bottle (i.e., determine differences between bottles or make the decision of which medication is needed). Thus, the ideal robot autonomy might fall somewhere in the middle of the taxonomy. To further consider which mid-autonomy category might be ideal, the designer can compare the possible effects of autonomy on robot-, human-, and social-related variables. Older adults were accepting of a robot delivering medication but less accepting of a robot determining which medication should be taken (Prakash et al. 2013). Thus, based on the human-related variable of acceptance, only some aspects of the task might be autonomously performed by the robot, whereas others should not. A robot autonomy level such as "batch processing" or "decision support" (where the user makes the final decision) may be suggested because it is important, in this example, to leave the decision-making to the human. Interestingly, as older adult opinions may differ for future generations (e.g., from cohort to cohort), or as reliable robot decision-making improves, then the autonomy level could be reconsidered and adjusted.

Now consider applying this framework to robot autonomy for a search and rescue operation. The first considerations are task and environment variables. Here, task criticality (high), and environmental complexity (high) pose challenges for a robot to reliably perform many aspects of the designated tasks. Thus, the initial recommendation may be to design the robot with low autonomy; a possible level would be "teleoperation." However, evaluation of human-robot performance (Riley & Endsley, 2004) during robot teleoperation in search and rescue environments suggests that SA is a challenge (e.g., determining which direction is up or down and what might happen next). Thus, a reconsideration of the autonomy level could affect an adjustment to "assisted teleoperation," whereby the robot could possibly provide some navigation cues, feedback, or override. A final consideration might be to design the robot so it has sliding autonomy, whereby it might shift autonomy modes (e.g., between teleoperation and assisted teleoperation), depending on the task complexity and operator's SA.

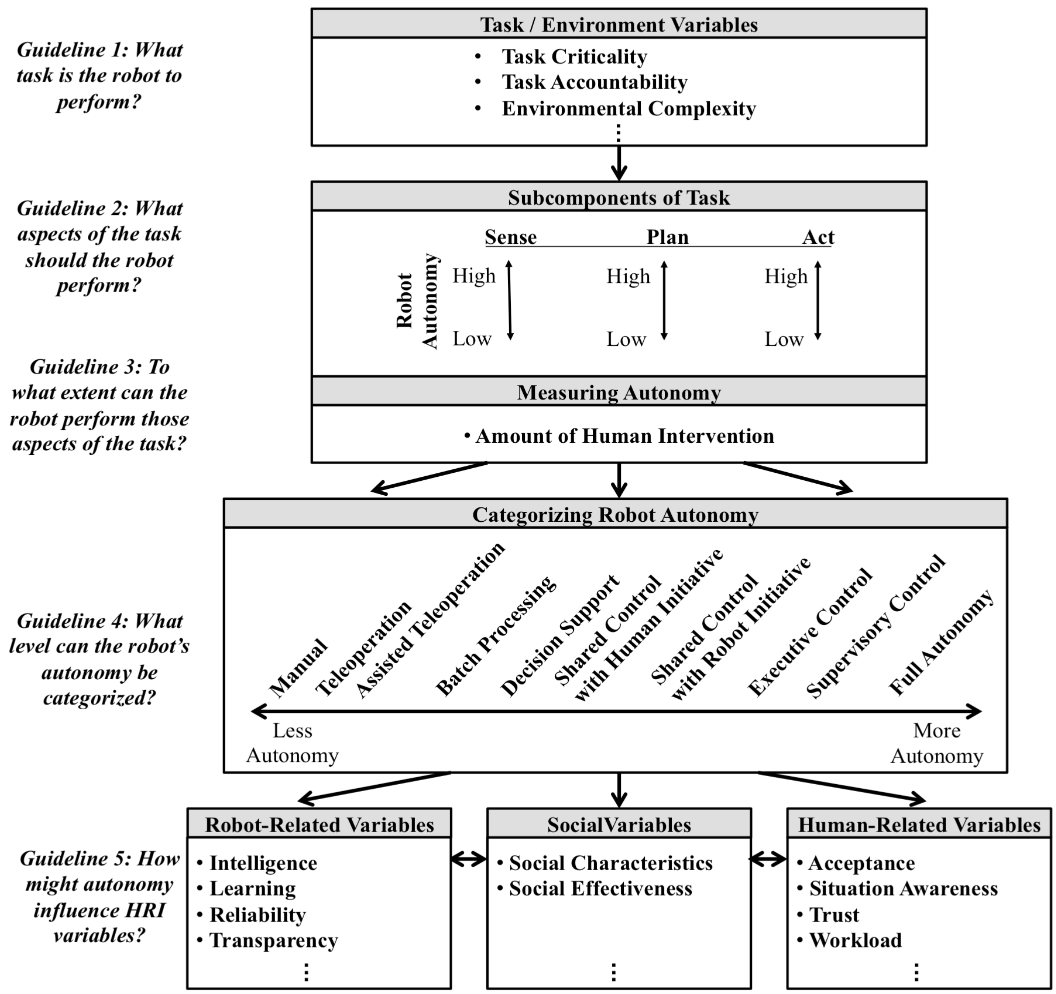

Figure 5. This framework can serve as (1) a set of guidelines suggesting task and environmental influences on robot autonomy, (2) guidelines for determining or measuring autonomy, (3) a taxonomy for categorizing autonomy, and finally, (4) a set of HRI variables that may be influenced by robot autonomy.

These thought experiments illustrate how the framework can serve as a guide to consider robot autonomy within the context of HRI. The advantage of categorizing a robot according to the taxonomy is to provide a meaningful and informative description of robot autonomy. Furthermore, consideration of how the robot's autonomy may influence HRI variables can assist designers and researchers in choosing an autonomy level that is appropriate for the task and user group. As these thought experiments demonstrate, the various guidelines may be considered and reconsidered throughout the design or evaluation process, making this framework a fluid guide rather than absolute method.

6. Conclusions

Levels of autonomy, ranging from teleoperation to fully autonomous systems, influence the nature of HRI. Our goal was to investigate robot autonomy within the context of HRI. We accomplished this by redefining the term *autonomy*, considering how the construct has been conceptualized within automation and HRI research. Our analysis led to the development of a framework for categorizing LORA and evaluating the effects of robot autonomy on HRI.

The framework provides a guide for appropriate selection of robot autonomy. The implementation of a service robot supplements a task but also changes human activity by imposing new demands on the human. Thus, the framework has scientific importance, beyond its use as a tool for guiding function allocation. As such, the framework conceptualizes autonomy along a continuum and identifies HRI variables that need to be evaluated as a function of robot autonomy. These variables include acceptance, SA, trust, robot intelligence, reliability, transparency, methods of control, and social interaction.

Many of the variables included in the framework require further research to better understand autonomy's complex role within HRI. HRI is a young field with substantial, albeit exciting, gaps in our understanding. Therefore, the proposed framework does not index causal relationships between variables and concepts. As the field of HRI develops, empirical research can be causally mapped to theoretical concepts and theories.

In summary, we have proposed a framework for LORA in HRI. Autonomy was defined within the context of HRI, and a taxonomy was proposed, not to provide exact descriptors of a robot's autonomy but rather to provide approximations of a robot's autonomy along a continuum. Additionally, guidelines were proposed to assist designers and researchers in identifying the appropriate LORA for any given task and potential influences on HRI. This framework is not meant to be used as a method, but as guidance for determining robot autonomy. A theme present in much of this investigation is that the role of autonomy in HRI is complex. Assigning a robot with an appropriate level of autonomy is important because a service robot changes human behavior. Implementing service robots has the potential to improve the quality of life for many people. At the same time, robot design will only be successful with consideration of how the robot's autonomy impacts HRI.

Acknowledgements

This research was supported in part by a grant from the National Institutes of Health (National Institute on Aging) Grant P01 AG17211 under the auspices of the Center for Research and Education on Aging and Technology Enhancement (CREATE; www.create-center.org). The authors would also like to thank Drs. Francis Durso, Charles Kemp, and Leila Takayama for their feedback on early versions of this work.

References

Alami, R., Chatila, R., Fleury, S., Ghallab, M., & Ingrand, F. (1998). An architecture for autonomy. *International Journal of Robotics Research, 17*(4), 315-337. doi: 10.1177/027836499801700402

Arkin, R. C. (1998). *Behavior Based Robotics*. Boston: MIT Press.

Baker, M., & Yanco, H. A. (2004). Autonomy mode suggestions for improving human-robot interaction. In *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics, 3,* 2948-2953. doi: 10.1109/ICSMC.2004.1400781

Beer, J. M., Fisk. A. D., & Rogers, W. A. (2012). Toward a psychological framework for levels of robot autonomy in human-robot interaction. *Technical Report HFA-TR-1204, Georgia Institute of Technology.* https://smartech.gatech.edu.

Bekey, G. A. (2005). *Autonomous robots: From biological inspiration to implementation and control*. Cambridge, MA: The MIT Press.

Breazeal, C. (2003). Emotion and sociable humanoid robots. *International Journal of Human Computer Interaction, 59*, 119-115. doi: 10.1016/S1071-5819(03)00018-1

Breazeal, C. (2005). Socially intelligent robots. *Interactions, 12*(2), 19-22. doi: 10.1145/1052438.1052455

Broadbent, E., Stafford, R. & MacDonald, B. (2009). Acceptance of healthcare robots for the older population: Review and future directions. *International Journal of Social Robotics, 1*(4), 319-330. doi: 10.1007/s12369-009-0030-6

Brooks, R. A. (2002). It's 2001 already. *Flesh and machines: How robots will change us* (63-98). New York, NY: Vintage Books, Random House Inc.

Bruemmer, D. J., Few, D. A., Boring, R. L., Marble, J. L., Walton, M. C., & Nielsen, C. W. (2005). Shared understanding for collaborative control. In *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics, 35*(4), 494-504. doi: 10.1109/TSMCA.2005.850599

Carlson, J., Murphy, R. R., & Nelson, A. (2004). Follow-up analysis of mobile robot failures. In *Proceedings of the IEEE International Conference on Robotics and Automation*, *5*, 4987-4994. New Orleans, Louisiana. doi: 10.1109/ROBOT.2004.1302508

Casper, J., & Murphy, R. (2003). Human-robot interactions during the robot-assisted urban search and rescue response at the World Trade Center. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 33*(3)*,* 367-385.

Cohen, M. S., Parasuraman, R., & Freeman, J. T. (1998). Trust in decision aids: A model and its training implications. In *Proceedings of the International Command and Control Research and Technology Symposium,* 1-37. Monterey, CA.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319-340. doi: 10.2307/249008

Desai, M., Kaniarasu, P., Medvedev, M., Steinfeld, A., & Yanco (2013). Impact of robot failures and feedback on real-time trust. In *Proceedings of the ACM/IEEE Conference on Human-Robot Interaction,* 251-258. Tokyo, Japan. doi: 10.1109/HRI.2013.6483596

Desai, M., Medvedev, M., Vazquez, M., McSheehy, S., Gadea-Omelchenko, S., Bruggeman, C., Steinfeld, A., & Yanco, H. (2012). Effects of changing reliability on trust of robot systems. In *Proceedings of the ACM/IEEE Conference on Human-Robot Interaction,* 73-80. doi: 10.1145/2157689.2157702

Desai, M., Stubbs, K., Steinfeld, A., & Yanco, H. (2009). Creating trustworthy robots: Lessons and inspirations from automated systems. In *Proceedings of the Society for the Study of Artificial Intelligence and the Simulation of Behaviour (AISB) Convention, New Frontiers in Human-Robot Interaction.*

Desai, M., & Yanco, H. A. (2005). Blending human and robot inputs for sliding scale autonomy. *The IEEE International Workshop on Robot and Human Interactive Communication*, 537-542. Nashville, Tennessee. doi: 10.1109/ROMAN.2005.1513835

Dewar, R. D., & Dutton, J. E. (1996). The adoption of radical and incremental innovations—An empirical- analysis. *Management Science, 32*(11), 1422-1433.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A. Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Computer Studies*, *58*, 697-718. doi: 10.1016/S1071-5819(03)00038-7

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*, 32-64. doi: 10.1518/001872095779049543

Endsley, M. R. (2006). Situation awareness. In G. Savendy (Ed.), *Handbook of human factors and ergonomics* (3rd ed.), pp. 528-542. New York, NY: Wiley.

Endsley, M. R., Bolte, B., & Jones, D. G. (2003). *Designing for situation awareness: An approach to human-centered design*. London: Taylor & Francis.

Endsley, M. R., & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics, 42*(3), 462-492. doi: 10.1080/001401399185595

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors, 37*(2), 381-394. doi: 10.1518/001872095779064555

Erikson, E. H. (1950). *Childhood and society*. New York, NY: W. W. Norton.

Ezer, N., Fisk, A. D., & Rogers, W. A. (2009). Attitudinal and intentional acceptance of domestic robots by younger and older adults. *Lecture Notes in Computer Science, 5615*, 39-48. doi: 10.1007/978-3-642-02710-9_5

Feil-Seifer, D., Skinner, K., & Mataric, M. J. (2007). Benchmarks for evaluating socially assistive robotics. *Interaction Studies*, *8*(3), 423-439. doi: 10.1075/is.8.3.07fei

Few, D., Smart, W. D., Bruemmer, D., & Neilsen, C. (2008). “Seamless autonomy”: Removing autonomy level stratifications. In *Proceedings of the Conference on Human System Interactions,* 446-451. Kraków, Poland.

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems, 42*, 143-166. doi: 10.1016/S0921-8890(02)00372-X

Franklin, S., & Graesser, A. (1996). Is it an agent, or just a program? A taxonomy for autonomous agents. In *Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages, Intelligent Agents*, 21-35. Budapest Hungary. doi: 10.1007/BFb0013570

Goodrich, M. A., & Olsen, D. R. (2003). Seven principles of efficient human robot interaction. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, 1-5*, 3943-3948. doi: 10.1109/ICSMC.2003.1244504

Goodrich, M. A., & Schultz, A. C. (2007). Human-robot interaction: A survey. *Foundations and Trends in Human-Computer Interaction, 1*(3), 203-275. doi: 10.1561/1100000005

Gorman, J. C., Cook, N. J., & Winner, J. L. (2006). Measuring team situation awareness in decentralized command and control environments. *Ergonomics, 49*(12-13), 1312-1325. doi: 10.1080/00140130600612788

Green, S. G., Gavin, M. B., & Aimansmith, L. (1995). Assessing a multidimensional measure of radical technological innovation. *IEEE Transactions on Engineering Management, 42*(3), 203-214. doi: 10.1109/17.403738

Groom, V., & Nass, C. (2007). Can robots be teammates? Benchmarks in human-robot teams. *Psychological Benchmarks of Human-Robot Interaction: Special issue of Interaction Studies, 8*(3), 483-500. doi: 10.1075/is.8.3.10gro

Hancock, P. A., Billings, D. R., & Schaefer, K. E. (2011). Can you trust your robot? *Ergonomics in Design, 19*(3), 24-29. doi: 10.1177/1064804611415045

Hearst, M. A. (1999). Mixed-initiative interaction: Trends and controversies. *IEEE Intelligent Systems*, *14*(5), 14-23.

Huang, H.-M, (2004). Autonomy levels for unmanned systems (ALFUS) framework volume I: Terminology version 1.1. In *Proceedings of the National Institute of Standards and Technology (NISTSP)*, Gaithersburg, MD.

Huang, H.-M, Messina, E. R., Wade, R. L., English, R. W, Novak, B., & Albus, J. S. (2004). Autonomy measures for robots. In *Proceedings of the International Mechanical Engineering Congress (IMECE),* 1-7. Anaheim, California. doi: 10.1115/IMECE2004-61812

Huang, H.-M, Pavek, K, Albus, J., & Messina, E. (2005). Autonomy levels for unmanned systems (ALFUS) framework: An update. In *Proceedings of the SPIE Defense and Security Symposium*, *5804*, 439-448. Orlando, Florida.

Huang, H.-M, Pavek, K., Novak, B., Albus, J. S., & Messina, E. (2005). A framework for autonomy levels for unmanned systems (ALFUS). In *Proceedings of the AUVSI's Unmanned Systems North America*, 849-863. Baltimore, Maryland.

Huang, H.-M, Pavek, K., Ragon, M., Jones, J., Messina, E., & Albus, J. (2007). Characterizing unmanned system autonomy: Contextual autonomous capability and level of autonomy analyses. In *Proceedings of the 2007 SPIE Defense and Security Symposium*. Orlando, Florida..

Johnson, M., Bradshaw, J. M., Feltovich, P. J., Jonker, C., Sierhuis, M., & van Riemsdijk, B. (2010). Toward Coactivity. In *Proceedings of the ACM/IEEE Conference on Human-Robot Interaction,* 101-102. Osaka, Japan. doi: 10.1145/1734454.1734492

Kaber, D. B., Onal, E., & Endsley, M. R. (2000). Design of automation for telerobots and the effect on performance, operator situation awareness, and subjective workload. *Human Factors and Ergonomics in Manufacturing, 10*(4), 409-430.

Kaber, D. B., Wright, M. C., & Sheik-Nainar, M. A. (2006). Investigation of multi-modal interface features for adaptive automation of a human-robot system. *International Journal of Human-Computer Studies, 64*, 527-540. doi: 10.1002/1520-6564(200023)10:4%3C409::AID-HFM4%3E3.0.CO;2-V

Kant, I. (1967). *Kant: Philosophical Correspondence, 1795-99.* (A. Zweig, Ed.). Chicago, IL: University of Chicago Press.

Khan, Z. (1998). Attitude towards intelligent service robots. *Numerical Analysis and Computer Science. Technical Report (TRITA-NA-P9821)*. Stockholm Sweden: Royal Institute of Technology.

Kim, T., & Hinds, P. (2006). Who should I blame? Effects of autonomy and transparency on attributions in human-robot interaction. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN),* 80-85. Hatfield, United Kingdom. doi: 10.1109/ROMAN.2006.314398

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*, 50-80. doi: 10.1518/hfes.46.1.50.30392

Lee, M. K. & Takayama, L. (2011). "Now I have a body": Uses and social norms of mobile remote presence in the workplace. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (SIGCHI),* 33-42. Vancouver, British Columbia. doi: 10.1145/1978942.1978950

Lewis, M., Wang, H., Chien, S. Y., Velagapudi, P., Scerri, P., & Sycara, K. (2010). Choosing autonomy modes for multi-robot search. *Human Factors, 52*(2), 225-233. doi: 10.1177/0018720810366859

Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human-human and human-automation trust: An integrative review. *Theoretical Issues in Ergonomics, 8*(4), 277-301. doi: 10.1080/14639220500337708

Milgram, P., Rastogi, A., & Grodski, J. J. (1995). Telerobotic control using augmented reality. *IEEE International Workshop on Robot and Human Communication,* 21-29. Tokyo, Japan. doi : 10.1109/ROMAN.1995.531930

Murphy, R. (2000). *Introduction to AI Robotics* (pp. 1-40). Cambridge, MA: The MIT Press.

Murphy, R. & Schreckenghost, D. (2013). Survey of metrics for human-robot interaction. In *Proceedings of the ACM/IEEE Conference on Human-Robot Interaction,* 197-198. Tokyo, Japan. doi: 10.1109/HRI.2013.6483569

Mutlu, B. & Forlizzi, J. (2008). Robots in organizations: Workflow, social, and environmental factors in human-robot interaction. In *Proceedings of the ACM/IEEE Conference on Human-Robot Interaction,* 239-248. Amsterdam, Netherlands. doi: 10.1145/1349822.1349860

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*(1), 81-103. doi: 10.1111/0022-4537.00153

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies, 45*(6), 669-678. doi: 10.1006/ijhc.1996.0073

Nass, C., Moon, Y., Fogg, B. J., & Reeves, B. (1995). Can computer personalities be human personalities? *International Journal of Human-Computer Studies, 43*(2), 223-239. doi: 10.1006/ijhc.1995.1042

Nass, C., Steuer, J., Henriksen, L., & Dryer, D. C. (1994). Machines, social attributions, and ethopoeia: Performance assessments of computers subsequent to 'self-' or 'other-' evaluations. *International Journal of Human-Computer Studies, 40*(3), 543-559. doi: 10.1006/ijhc.1994.1025

Olsen, D. R., & Goodrich, M. A. (2003). Metrics for evaluating human-robot interactions. In *Proceedings of NIST Performance Metrics for Intelligent Systems Workshop*. Gaithersburg, Maryland.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*(2), 230-253. doi: 10.1518/001872097778543886

Parasuraman, R., & Wickens, C. D. (2008). Humans: Still vital after all these years of automation. *Human Factors, 50*(3), 511-520. doi: 10.1518/001872008X312198

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems Man and Cybernetics Part A: Systems and Humans, 30*(3), 286-297. doi: 10.1109/3468.844354

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making, 2*(2), 140-160. doi: 10.1518/155534308X284417

Piaget, J. (1932). *The Moral Judgment of a Child*. Glencoe, IL: The Free Press.

Prakash, A., Beer, J. M., Deyle, T., Smarr, C.-A., Che, T. L., Mitzner, T. L., Kemp, D. C., & Rogers, W. A. (2013). Older adults' medication management in the home: How can robots help? In *Proceedings of the ACM/IEEE Conference on Human-Robot Interaction,* 283-290. Tokyo, Japan. doi: 10.1109/HRI.2013.6483600

Rosen, C. A. & Nilsson, N. J (1966). *Application Of Intelligent Automata to Reconnaissance*. Technical Report. Menlo Park, CA: Stanford Research Institute.

Riley, J. M., & Endsley, M. R. (2004). The hunt for situation awareness: Human-robot interaction in search and rescue. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting,* 693-697. New Orleans, Louisiana.doi: 10.1177/154193120404800389

Russell, S. J., & Norvig, P. (2003). *Artificial intelligence: A modern approach* (2nd ed.). Upper Saddle River, NJ: Pearson Education, Inc.

Scholtz, J. (2002a). Theory and evaluation of human-robot interactions. In *Proceedings of the IEEE International Conference on System Sciences*. Waikoloa Village, Hawaii. doi: 10.1109/HICSS.2003.1174284

Scholtz, J. (2002b). Human-robot interactions: Creating synergistic cyber forces. In *Proceedings from the NRL Workshop on Multi-Robot Systems: From Swarms to Intelligent Automata.* 177-184. Washington, DC. doi: 10.1007/978-94-017-2376-3_19

Scholtz, J., Antonishek, B., & Young, J. (2004). Evaluation of a human-robot interface: Development of a situational awareness methodology. In *Proceedings of the International Conference on System Sciences*, 1-9. Waikoloa Village, Hawaii. doi: 10.1109/HICSS.2004.1265327

Sheridan, T. B., & Verplank, W. L. (1978). Human and computer control of undersea teleoperators. *Man-Machine Systems Laboratory Report*. Cambridge, MA: MIT.

Skinner, B. F. (1978). *Reflection on behaviorism and society*. Englewood Cliffs, NJ: Prentice-Hall.

Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., & Goodrich, M. (2006). Common metrics for human-robot interaction. In *Proceedings of Human-Robot Interaction Conference*, 33-40. Salt Lake City, Utah. doi: 10.1145/1121241.1121249

Stubbs, K., Hinds, P., & Wettergreen, D. (2007). Autonomy and common ground in human-robot interaction: A field study. *IEEE Intelligent Systems: Special Issue on Interacting with Autonomy*, *22*(2), 42-50. doi: 10.1109/MIS.2007.21

Takayama, L., Marder-Eppstein, E., Harris, H., & Beer, J. M. (2011). Assisted driving of a mobile remote presence system: System design and controlled user evaluation. In *Proceedings of the International Conference on Robotics and Automation (ICRA)*, 1883-1889. Shanghai, CN. doi: 10.1109/ICRA.2011.5979637

Tarn, T.-J., Zi, N., Guo, C., & Bejczy, A. K. (1995). Function-based control sharing for robotics systems. In *Proceedings on IEEE International Conference on Intelligent Robots and Systems, 3*, 1-6. Pittsburgh, Pennsylvania.

Thrun, S. (2004). Toward a framework for human-robot interaction. *Human-Computer Interaction, 19*(1-2), 9-24. doi: 10.1207/s15327051hci1901&2_2

Tiwari, P., Warren, J., Day, K. J., & MacDonald, B. (2009). Some non-technology implications for wider application of robots to assist older people. In *Proceedings of the Conference and Exhibition of Health Informatics.* New Zealand.

Tsui, K. M., Desai, M., Yanco, H., & Uhlik, C. (2011). Exploring use cases for telepresence robots. In *Proceedings of the ACM/IEEE Conference on Human-Robot Interaction,* 11-18. *Lausanne, Switzerland.* doi: 10.1145/1957656.1957664

Tsang, P. S. & Vidulich, M. A. (2006). Mental workload and situation awareness. In G. Savendy (Ed.), *Handbook of Human Factors and Ergonomics* (3rd ed., pp. 243-268). New York, NY: Wiley. doi: 10.1002/0470048204.ch9

Urdiales, C., Poncela, A., Sanchez-Tato, I., Galluppi, F., Olivetti, M., & Sandoval, F. (2007). Efficiency based reactive shared control for collaborative human/robot navigation. In *Proceedings from the IEEE/RSJ International Conference on Intelligent Robots and Systems* (IROS), 3586-3591. San Diego, California. doi: 10.1109/IROS.2007.4399288

Wellman, H. (1992). *The Child's Theory of Mind*. Boston, MA: MIT Press.

Wooldridge, M., & Jennings, N. R. (1995). Intelligent agents: Theory and practice. *Knowledge Engineering Review, 10,* 115-152. doi: 10.1017/S0269888900008122

Yanco, H. & Drury, J. (2004). Classifying human-robot interaction: An updated taxonomy. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, 3*, 2841-2846. Hague, Netherlands. doi: 10.1109/ICSMC.2004.1400763

Young, J., Hawkins, R., Sharlin, E., and Igarashi, T. (2009). Toward acceptable domestic robots: Applying insights from social psychology. *International Journal of Social Robotics*, *1*(1), 95-108. doi: 10.1007/s12369-008-0006-y

---

Jenay M. Beer, Assistive Robotics and Technology Lab; Department of Computer Science and Engineering; University of South Carolina; 315 Main St, Columbia, SC 29208; Email: jbeer@cse.sc.edu