



Published in final edited form as:

J Hum Robot Interact. 2014 July ; 3(2): 74–99. doi:10.5898/JHRI.3.2.Beer.

Toward a framework for levels of robot autonomy in human-robot interaction

Jenay M. Beer,

School of Psychology, Georgia Institute of Technology

Arthur D. Fisk, and

School of Psychology, Georgia Institute of Technology

Wendy A. Rogers

School of Psychology, Georgia Institute of Technology

Abstract

A critical construct related to human-robot interaction (HRI) is autonomy, which 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 may 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 levels of robot autonomy in HRI. To reach this goal, the 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, the 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 guidelines to determine autonomy, categorize the LORA along a qualitative taxonomy, and consider which HRI variables (e.g., acceptance, situation awareness, reliability) may be influenced by the LORA.

Keywords

Human-robot interaction; automation; autonomy; levels of robot autonomy; framework

Introduction

Autonomy: from Greek *autos* (“self,”) and *nomos* (“law”)

“I am putting myself to the fullest possible use...”

–HAL 9000 (2001: Space Odyssey)

Developing fully autonomous robots has been a goal of roboticists and other visionaries since the emergence of the field, both in product development and science fiction. However, 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. 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, and to what extent.” A

scientific base of empirical research can guide designers in identifying appropriate trade-offs 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 human-robot interaction.

The field of human-robot interaction (HRI) largely lacks frameworks and conceptual models that organize empirical observations to theoretical concepts. 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 class of robots is broad, there are certain characteristics 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 may be expected to interact with humans with limited or no formal training (Thrun, 2004).

Goals and contributions

Our overarching goal was two-fold: parse the literature to better understand autonomy and to critique autonomy from a psychological perspective and identify variables that influence, and are influenced by, autonomy. Specifically, the objectives of this paper were to: 1) Refine the definition of autonomy; 2) Propose a process of determining levels of robot autonomy; 3) Suggest a framework to identify potential variables related to autonomy.

Autonomy defined

Autonomy has been of both philosophical and psychological interest for over 300 years. In the 18th century, autonomy was most famously considered by philosopher Immanuel 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. However in psychology, autonomy has been primarily discussed in relation to child development. In that literature, the term autonomy is discussed as a subjective construct involving self-control, governing, and free will. For instance, Piaget (1932) proposed that autonomy is the ability to self govern, and a critical component in a child's moral development. Erikson (1997) similarly defined autonomy as a child's development of a sense of self control (e.g., early childhood toilet training). Children who successfully develop autonomy feel secure and confident; children who do not develop autonomy may experience self-doubt and shame.

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 technology, particularly 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 human intervention. The various definitions of robot autonomy are presented in Table 1.

To clarify the term autonomy we propose the following. A weak (i.e. global) definition of autonomy is: *the extent to which a system can carry out its own processes and operations without external control*. This weak definition of autonomy can be used to denote autonomous capabilities of humans or machines. However, a stronger more specific definition can be given to robots, by integrating the definitions provided in Table 1. Autonomy, as related to robots, we define as:

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

The proposed strong definition of autonomy integrates the current definitions of autonomy, and highlights the prevalent characteristics of autonomy (i.e., sense, plan, act, goal, and control). Note that both the weak and strong definition begin with the phrase “to the extent to which...” This choice in wording exemplifies that autonomy is not all or nothing. Autonomy exists on a continuum from no autonomy to full autonomy.

Autonomy in automation

To guide our investigation of autonomy in HRI, we first reviewed the human-automation literature. Automation researchers have a history of studying and understanding human-related variables, and much of this literature can be informative to 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). However, 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.

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 levels of automation 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 and action selection. It lacked specification of *input* functions related to information acquisition (i.e., sensing) or the processing of that information (i.e., formulating options or strategies).

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 strength of the Endsley and Kaber model is the detail used to describe each of the automation levels. The taxonomy is 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 chosen option.

Parasuraman, Sheridan, and Wickens (2000) proposed the most recent model for types and levels of automation (Figure 1). 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*.

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., in air traffic control an automated 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 the automatic stapler in a photocopy machine, or autopilot in an aircraft.

The bottom of the flow chart in Figure 1 depicts primary and secondary evaluative criteria. These evaluative criteria were meant to provide a guide for determining a system's level of automation. 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. Therefore, primary evaluative criteria are evaluated (e.g., workload, situation awareness), and then the level of automation is adjusted. Next secondary criteria are evaluated (e.g., automation reliability, cost of action outcomes), and again the level of automation is adjusted. This iterative process provides a starting point for determining appropriate levels of automation to be implemented in a particular system.

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 the Sheridan and Verplank's taxonomy has been suggested as appropriate to describe how autonomous a robot is (Goodrich & Schultz, 2007). However, it is important to consider the differences between automation and robotics. Robots may serve different functions relative to traditional automation; for example, some 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.

Autonomy in HRI

Autonomy within an HRI context is a widely considered construct; however, the ideas surrounding how autonomy influences human-robot interaction are varied. There are two schools of thought: (1) higher robot autonomy requires *lower levels or less frequent HRI*; and (2) higher robot autonomy requires *higher levels or more sophisticated forms of HRI*.

The first viewpoint, that higher autonomy requires less HRI, was proposed by Huang and colleagues (Huang, Messina, Wade, English, Novak, & Albus, 2004; Huang, Pavek, Albus, & Messina, 2005; Huang, Pavek, Novak, Albus & Messina, 2005; Huang, Pavek, Ragon, Jones, Messina, & Albus, 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 as a basis for HRI autonomy classes more generally (Yanco & Drury, 2004).

The Huang framework stated that the relationship between the level of human-robot interaction 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 HRI so that as the level of robot autonomy increases the HRI decreases (see Figure 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-robot ratio.

Other researchers have also proposed that higher robot autonomy requires less interaction (Yanco & Drury, 2004). Autonomy has been described as the amount that a person can neglect the robot; 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 can be neglected for a longer time period.

"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 human-robot interaction is a stark contrast to the way in which other HRI researchers consider autonomy; they have proposed that more robot autonomy requires more human-robot 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) mandated higher degrees of autonomy because they operate in a variable environment and interact in close proximity to people. Thrun declared that “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” (2004, p. 14).

Furthermore, autonomy has been proposed as a benchmark for developing *social interaction* in socially assistive robotics (Feil-Seifer, Skinner, & Mataric, 2007). They proposed that autonomy serves two functions: to perform well in a desired task, and to be proactively social. However, they warned that the robot’s autonomy should allow for social interaction only when appropriate (i.e., only when social interaction enhances performance). However, 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).

The seemingly dichotomous HRI viewpoints may be due in part to inconsistent use of terminology. There is ambiguity concerning the precise meaning of the “I” in HRI; *intervention* and *interaction* have been used synonymously. The ambiguous use of these terms makes unclear how autonomy should be measured. Conceivably intervention could be interpreted as a specific type of interaction (as suggested in Huang et al., 2004). The goal of some researchers for a robot to act autonomously with no HRI mirrors the human *out of the loop* phenomenon in automation, which is known to cause performance problems (e.g., Endsley, 2006; Endsley & Kiris, 1995).

A framework of autonomy and HRI is needed. As this literature review revealed, autonomy is an important construct related to HRI, and a multi-disciplinary approach to developing a framework is essential. Now that a definition of autonomy has been established, and inconsistencies in the literature identified, we move toward developing the building blocks for a framework of robot autonomy.

Toward a Framework of Levels of Robot Autonomy and HRI—We provide a framework for examining levels of robot autonomy and its effect on human-robot interaction. This framework can be used as an organizing flow chart, consisting of several stages (Figure 3). Stages 1–3 serve as a guideline to determine robot autonomy. Stage 4 categorizes robot autonomy via a 10-point taxonomy. Finally, Stage 5 broadly suggests the

implications of the robot autonomy on HRI (i.e., human variables, robot variables, and interaction variables), and a conceptual model of the framework is presented.

Determining Robot Autonomy (Stages 1–3)

Stages 1–3 are meant to provide guidelines for determining and measuring robot autonomy. Specifically, the proposed guidelines in this section focus on human-robot interaction, 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 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 floor, cleaning filters, emptying dirt bin/bag) 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), the Roomba's autonomy could be categorized differently, as it is currently incapable of vacuuming stairs. Therefore, specifying the context of the task and environment is critical for determining the task-specific level of robot autonomy.

Secondly, specifying demands, such as task criticality, accountability, 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). For example, in many cases failures or errors at early stages of automation are not as critical as errors at later stages of automation. One rational 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 human safety (Parasuraman, Sheridan, & Wickens, 2000; Parasuraman & Wickens, 2008). For example, unreliability in a robot that autonomously navigates may results 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 determines what medication a patient should take. In this example, robot failure may result in critical, if not lethal, consequences. Responsibility for the success of task completion (i.e., accountability) 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 accountability may be perceived to be split between the robot and 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 shall perform, as well as in designing the system so human supervisors are held accountable and can easily diagnose and alleviate consequences of error.

Furthermore, environmental complexity is also a critical demand to consider when defining autonomy. Service robot designed for assistive functions (e.g., home or healthcare applications), surveillance, or first responders (e.g., search and rescue) will be required to operate in unknown, unstructured, and dynamic environments. Functioning in such difficult environments 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 sensors to perceive the world around it. Higher levels of robot autonomy may be required for a service robot to function in unstructured ever-changing environments (Thrun, 2004). That is, the robot must have the autonomy to make changes in goals and actions based on the sensor data of the dynamic environment. 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 determining 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). 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), relatively high level autonomy in PLAN (except the human provided the high level goal), and a high level of autonomy in physically implementing the plan (ACT).

As these examples suggest, autonomy can vary along any of the SENSE, PLAN, and ACT primitives, which relates to the next determining question:

“To what extent can the robot perform those 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).

As depicted in Figure 4 the level of autonomy may vary from low to high for each of the robot primitives. Determining the robot autonomy prompts a clarification of how to *measure* the extent or degree to which a robot can perform each of those aspects (SENSE, PLAN, ACT) of the task. In the automation literature, level of autonomy is most often identified by function allocation. Consider the Endsley and Kaber (1999) model, the level of automation is specified in their taxonomy based on the allocation of function to either the human or automation. For instance, in their automation level Automated Decision Making: the automation selects and carries out an option; the human can have input in the alternatives generated by the automation.

In HRI the allocation of function 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%. For example, a teleoperated robot has autonomy=0%, and intervention=100%. A fully autonomous robot has autonomy=100%, and intervention=0%. 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, autonomy=75% and intervention=25%. 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 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/control method, and response time of the user. Therefore, if interaction *time* is used as a quantitative measure, care should be taken when generalizing those findings to other robot systems or tasks. 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, rather than 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 indication of the robot's degree of autonomy.

Intervention is defined as the human performing some aspect of the task. As we have discussed earlier, *intervention* and *interaction* are not necessarily interchangeable terms. Intervention is a type of interaction specific to task sharing. Interaction may include other factors not necessarily specific to the intervention of task completion, such as verbal communication, gestures, or emotion 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, which 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.

Categorizing Levels of Robot Autonomy (LORA) for HRI: A Taxonomy (Stage 4)

Once guidelines for determining robot autonomy have been followed, the next stage is to categorize the robot's autonomy along a continuum. Specification of intermediate autonomy levels is a limitation in previous HRI frameworks (e.g., Yanco & Drury, 2004; Huang, Pavek, Albus, & Messina, 2005). In Table 4 we propose a taxonomy in which the robot autonomy can be categorized into "levels of robot autonomy" (LORA).

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 each play. That is, for each proposed LORA, we suggest the (1) function allocation between robot and human for each of the SENSE, PLAN, ACT primitives; (2) provide a proposed description of each LORA; and (3) illustrate with examples of service robots from the HRI literature. Table 4 includes a mix of empirical studies involving robots and simulations, as well as published robot autonomy architectures. Autonomy is a continuum with blurred borders between the proposed levels. The levels should not be treated as exact descriptors of a robot's autonomy but rather treated as an approximation of a robot's autonomy level along the continuum.

The Influence of Autonomy on HRI (Stage 5)

The last stage of the framework is 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, the robot, and the interaction between the two (see Beer, Fisk, and Rogers (2012) for a more complete literature review of the variables listed). 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 need further investigation. Evaluation of the interaction between autonomy and HRI related variables can be used as evaluative criteria to determine if the autonomy level of the robot is appropriate. Thus, the framework can be a tool for guiding appropriate autonomy levels that support optimal human-robot interaction.

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 programs, or cognitive functions. 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; Russel & Norvig, 2003; Thrun, 2003).

As robots move from the laboratory to more dynamic environments (e.g., the home, hospital setting, the 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 the machine will sometimes fail and consideration for how such failures will consequence human performance. In automation, degraded reliability at higher levels of autonomy resulted in an “out of the loop” operator (Endsley, 2006), 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 observe the system and understand what it is doing. Automated tasks where an operator can form a mental model are referred to as transparent. 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.

Human-related variables—Situation awareness (SA) and mental workload have a long history in the automation literature. These concepts that are intricately 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 level of automation (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 their own 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, however, studies that directly investigate the effect of autonomy on trust have not been conducted. Nonetheless, 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.

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 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 (i.e., "adjustable autonomy"). 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 (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 research is needed to understand and model the variables that influence robot acceptance, how such variables interact, and finally how predictive value varies over the autonomy continuum.

Interaction-related variables—Understanding social interaction in HRI and its relation to autonomy has been a topic of science fiction, media, and film for decades. In fact, 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 behave, then it becomes all the more important to understand how match user expectations with the robot's actual autonomy. According to Breazeal (2003), when designing robots, the emphasis should not be whether people will develop a social model to understand robots. Rather, it is more important that the robot adhere to the social models the humans expect. What social models do people hold for robots? And do those social models change as a function of robot autonomy?

It is accepted in the research community 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 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 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 human-robot interaction.

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 to 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, social interaction may not exist between the robot and human, *per se*. Rather, the robot may be designed to facilitate social communication *between people* (i.e., the operator and a remotely located individual). 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/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).

A Framework of Robot Autonomy

The framework of robot autonomy (stages 1–5) is depicted in Figure 5. From top to bottom, the model depicts the guideline stages. By no means should this model be treated as a method. Rather the framework and taxonomy should be treated as guidelines to determine autonomy, categorize the LORA along a qualitative scale, and consider which HRI variables may be influenced by the LORA.

Conclusion

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 the 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 in 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 levels of robot autonomy in human-robot interaction. Autonomy is conceptualized as a continuum from teleoperation to full autonomy. A 10-point taxonomy was proposed, not as exact descriptors of a robot's autonomy, but rather to provide approximations of a robot's autonomy along the continuum. Additionally, a conceptual model was developed to guide researchers in identifying the appropriate level of robot autonomy for any given task and potential influences on human-robot interaction. 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. But robot design will only be successful with consideration of how the robot's autonomy will impact the human-robot interaction.

Acknowledgments

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 Francis Durso and Charles Kemp from Georgia Institute of Technology, as well as Leila Takayama from Willow Garage (www.willowgarage.com) for their feedback on early versions of this work.

References

- Alami R, Chatila R, Fleury S, Ghallab M, Ingrand F. An architecture for autonomy. *International Journal of Robotics Research*. 1998; 17(4):315–337.
- Arkin, RC. *Behavior Based Robotics*. Boston: MIT Press; 1998.

- Baker M, Yanco HA. Autonomy mode suggestions for improving human-robot interaction. *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics*. 2004; 3:2948–2953.
- Bekey, GA. *Autonomous Robots: From Biological Inspiration to Implementation and Control*. Cambridge, MA: The MIT Press; 2005.
- Breazeal C. Emotion and Sociable Humanoid Robots. *International Journal of Human Computer Interaction*. 2003; 59:119–115.
- Breazeal C. Socially intelligent robots. *Interactions*. 2005; 12(2):19–22.
- Broadbent E, Stafford R, MacDonald B. Acceptance of healthcare robots for the older population: Review and future directions. *International Journal of Social Robotics*. 2009; 1(4):319–330.
- Brooks, RA. *Flesh and Machines: How Robots will Change Us*. New York: Vintage Books, Random House Inc; 2002. It's 2001 already; p. 63-98.
- Bruemmer DJ, Few DA, Boring RL, Marble JL, Walton MC, Nielsen CW. Shared understanding for collaborative control. *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics*. 2005; 35(4):494–504.
- Carlson, J., Murphy, RR., Nelson, A. Follow-up analysis of mobile robot failures. *Proceedings of the IEEE International Conference on Robotics and Automation*; 2004. p. 4987-4994.
- Cohen, MS., Parasuraman, R., Freeman, JT. Trust in decision aids: A model and its training implications. *Proceedings of the International Command and Control Research and Technology Symposium*; Monterey, CA. 1998. p. 1-37.
- Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*. 1989; 13(3):319–340.
- Desai, M., Yanco, HA. Blending human and robot inputs for sliding scale autonomy. *The IEEE International Workshop on Robot and Human Interactive Communication*; 2005. p. 537-542.
- Desai, M., Stubbs, K., Steinfeld, A., Yanco, H. Creating trustworthy robots: Lessons and inspirations from automated systems. *Proceedings of the AISB Convention, New Frontiers in Human-Robot Interaction*; 2009.
- Dewar RD, Dutton JE. The adoption of radical and incremental innovations – an empirical-analysis. *Management Science*. 1996; 32(11):1422–1433.
- Dzindolet MT, Peterson SA, Pomranky RA, Pierce LG, Beck HP. The role of trust in automation reliance. *International Journal of Computer Studies*. 2003; 58:697–718.
- Endsley MR. Toward a theory of situation awareness in dynamic systems. *Human Factors*. 1995; 37:32–64.
- Endsley, MR. Situation Awareness. In: Savendy, G., editor. *Handbook of Human Factors and Ergonomics*. 3. New York: Wiley; 2006. p. 528-542.
- Endsley, MR., Bolte, B., Jones, DG. *Designing for Situation Awareness: An Approach to Human-Centered Design*. London: Taylor & Francis; 2003.
- Endsley MR, Kaber DB. Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*. 1999; 42(3):462–492. [PubMed: 10048306]
- Endsley MR, Kiris EO. The out-of-the-loop performance problem and level of control in automation. *Human Factors*. 1995; 37(2):381–394.
- Erikson, EH. *The Life Cycle Completed*. New York, N.Y: Norton; 1997.
- Ezer N, Fisk AD, Rogers WA. Attitudinal and intentional acceptance of domestic robots by younger and older adults. *Lecture notes in computer science*. 2009; 5615:39–48.
- Feil-Seifer D, Skinner K, Mataric MJ. Benchmarks for evaluating socially assistive robotics. *Interaction Studies*. 2007; 8(3):423–439.
- Few, D., Smart, WD., Bruemmer, D., Neilsen, C. “Seamless autonomy”: Removing autonomy level stratifications. *Proceedings of the Conference on Human System Interactions*; 2008. p. 446-451.
- Fong T, Nourbakhsh I, Dautenhahn K. A survey of socially interactive robots. *Robotics and Autonomous Systems*. 2003; 42:143–166.
- Franklin, S., Graesser, A. Is it an agent, or just a program? A taxonomy for autonomous agents. *Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages, Intelligent Agents III*; 1997. p. 21-35.

- Goodrich, MA., Olsen, DR. Seven principles of efficient human robot interaction. Proceedings of the IEEE International Conference on Systems, Man and Cybernetics; 2003. p. 1-5.p. 3943-3948.
- Goodrich MA, Schultz AC. Human-robot interaction: A survey. *Foundations and Trends in Human-Computer Interaction*. 2007; 1(3):203–275.
- Gorman JC, Cook NJ, Winner JL. Measuring team situation awareness in decentralized command and control environments. *Ergonomics*. 2006; 49(12–13):1312–1325. [PubMed: 17008258]
- Green SG, Gavin MB, Aïmansmith L. Assessing a multidimensional measure of radical technological innovation. *IEEE Transactions on Engineering Management*. 1995; 42(3):203–214.
- Groom V, Nass C. Can robots be teammates? Benchmarks in human-robot teams. *Psychological Benchmarks of Human-Robot Interaction: Special issue of Interaction Studies*. 2007; 8(3):483–500.
- Hancock PA, Billings DR, Schaefer KE. Can you trust your robot? *Ergonomics in Design*. 2011; 19(3): 24–29.
- Hearst, MA. IEEE Intelligent Systems. 1999. Mixed-initiative interaction: Trends and controversies; p. 14-23.
- Huang, H-M. Autonomy levels for unmanned systems (ALFUS) framework volume I: Terminology version 1.1. Proceedings of the National Institute of Standards and Technology (NISTSP); Gaithersburg, MD. 2004.
- Huang, H-M., Messina, ER., Wade, RL., English, RW., Novak, B., Albus, JS. Autonomy measures for robots. Proceedings of the International Mechanical Engineering Congress (IMECE); 2004. p. 1-7.
- Huang H-M, Pavek K, Albus J, Messina E. Autonomy levels for unmanned systems (ALFUS) framework: An update. Proceedings of the SPIE Defense and Security Symposium. 2005; 5804:439–448.
- Huang, H-M., Pavek, K., Novak, B., Albus, JS., Messina, E. A framework for autonomy levels for unmanned systems (ALFUS). Proceedings of the AUVSI's Unmanned Systems North America; Baltimore, Maryland. 2005. p. 849-863.
- Huang, H-M., Pavek, K., Ragon, M., Jones, J., Messina, E., Albus, J. The Unmanned Systems Technology IX, The International Society for Optical Engineering. 2007. Characterizing unmanned system autonomy: Contextual autonomous capability and level of autonomy analyses; p. 6561
- Kaber DB, Onal E, Endsley MR. Design of automation for telerobots and the effect on performance, operator situation awareness, and subjective workload. *Human Factors and Ergonomics in Manufacturing*. 2000; 10(4):409–430.
- Kaber DB, Wright MC, Sheik-Nainar MA. Investigation of multi-modal interface features for adaptive automation of a human-robot system. *International Journal of Human-Computer Studies*. 2006; 64:527–540.
- Kant, I. Kant: Philosophical Correspondence, 1795–99. Zweig, A., editor. Chicago: University of Chicago Press; 1967.
- Khan, Z. Numerical Analysis and Computer Science Tech.. Rep. (TRITA-NA-P9821). Stockholm Sweden: Royal Institute of Technology; 1998. Attitude towards intelligent service robots.
- Kim, T., Hinds, P. Who should I blame? Effects of autonomy and transparency on attributions in human-robot interaction. Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN); 2006. p. 80-85.
- Lee JD, See KA. Trust in automation: Designing for appropriate reliance. *Human Factors*. 2004; 46:50–80. [PubMed: 15151155]
- Madhavan P, Wiegmann DA. Similarities and differences between human-human and human-automation trust: An integrative review. *Theoretical Issues in Ergonomics*. 2007; 8(4):277–301.
- Milgram, P., Rastogi, A., Grodski, JJ. Telerobotic control using augmented reality. *IEEE International Workshop on Robot and Human Communication*; 1995. p. 21-29.
- Murphy, RR. Introduction to AI Robotics. Cambridge, MA: The MIT Press; 2000. p. 1-40.
- Nass C, Moon Y. Machines and mindlessness: Social responses to computers. *Journal of Social Issues*. 2000; 56(1):81–103.

- Nass C, Fogg BJ, Moon Y. Can computers be teammates? *International Journal of Human-Computer Studies*. 1996; 45(6):669–678.
- Nass C, Moon Y, Fogg BJ, Reeves B. Can computer personalities be human personalities? *International Journal of Human-Computer Studies*. 1995; 43(2):223–239.
- Nass C, Steuer J, Henriksen L, Dryer DC. Machines, social attributions, and ethopoeia: Performance assessments of computers subsequent to ‘self-’ or ‘other-’ evaluations. *International Journal of Human-Computer Studies*. 1994; 40(3):543–559.
- Olsen, DR., Goodrich, MA. Metrics for evaluating human-robot interactions. *Proceedings of NIST Performance Metrics for Intelligent Systems Workshop*; 2003.
- Parasuraman R, Riley V. Humans and automation: Use, misuse, disuse, abuse. *Human Factors*. 1997; 39(2):230–253.
- Parasuraman R, Wickens CD. Humans: Still vital after all these years of automation. *Human Factors*. 2008; 50(3):511–520. [PubMed: 18689061]
- Parasuraman R, Sheridan TB, Wickens CD. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems Man and Cybernetics Part A: Systems and Humans*. 2000; 30(3):286–297.
- Parasuraman R, Sheridan TB, Wickens CD. Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making*. 2008; 2(2):140–160.
- Piaget, J. *The Moral Judgment of a Child*. Glencoe, IL: The Free Press; 1932.
- Riley, JM., Endsley, MR. The hunt for situation awareness: Human-robot interaction in search and rescue. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*; 2004. p. 693-697.
- Russell, SJ., Norvig, P. *Artificial Intelligence: A Modern Approach*. 2. Upper Saddle River, NJ: Pearson Education, Inc; 2003.
- Scholtz, J., Antonishek, B., Young, J. Evaluation of a human-robot interface: Development of a situational awareness methodology. *Proceedings of the International Conference on System Sciences*; 2004. p. 1-9.
- Sheridan, TB., Verplank, WL. *Human and computer control of undersea teleoperators (Man-Machine Systems Laboratory Report)*. Cambridge: MIT; 1978.
- Skinner, BF. *Reflection on Behaviorism and Society*. Englewood Cliffs, N.J: Prentice-Hall; 1978.
- Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., Goodrich, M. Common metrics for human-robot interaction. *Proceedings of Human-Robot Interaction Conference*; 2006. p. 33-40.
- Takayama, L., Marder-Eppstein, E., Harris, H., Beer, JM. Assisted driving of a mobile remote presence system: System design and controlled user evaluation. *Proceedings of the International Conference on Robotics and Automation: ICRA; Shanghai, CN*. 2011. p. 1883-1889.
- Tarn TJ, Zi N, Guo C, Bejczy AK. Function-based control sharing for robotics systems. *Proceedings on IEEE International Conference on Intelligent Robots and Systems*. 1995; 3:1–6.
- Thrun S. Toward a framework for human-robot interaction. *Human-Computer Interaction*. 2004; 19(1–2):9–24.
- Tiwari, P., Warren, J., Day, KJ., MacDonald, B. Some non-technology implications for wider application of robots to assist older people. *Proceedings of the Conference and Exhibition of Health Informatics; New Zealand*. 2009.
- Tsang, PS., Vidulich, MA. Mental workload and situation awareness. In: Savendy, G., editor. *Handbook of Human Factors and Ergonomics*. 3. New York: Wiley; 2006. p. 243-268.
- Urdiales, C., Poncela, A., Sanchez-Tato, I., Galluppi, F., Olivetti, M., Sandoval, F. Efficiency based reactive shared control for collaborative human/robot navigation. *Proceedings from the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*; 2007. p. 3586-3591.
- Wooldridge M, Jennings NR. Intelligent agents: Theory and practice. *Knowledge Engineering Review*. 1995; 10:115–152.
- Yanco H, Drury J. Classifying human-robot interaction: an updated taxonomy. *IEEE International Conference on Systems, Man and Cybernetics*. 2004; 3:2841–2846.

Young J, Hawkins R, Sharlin E, Igarashi T. Toward acceptable domestic robots: Applying insights from social psychology. *International Journal of Social Robotics*. 2009; 1(1):95–108.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

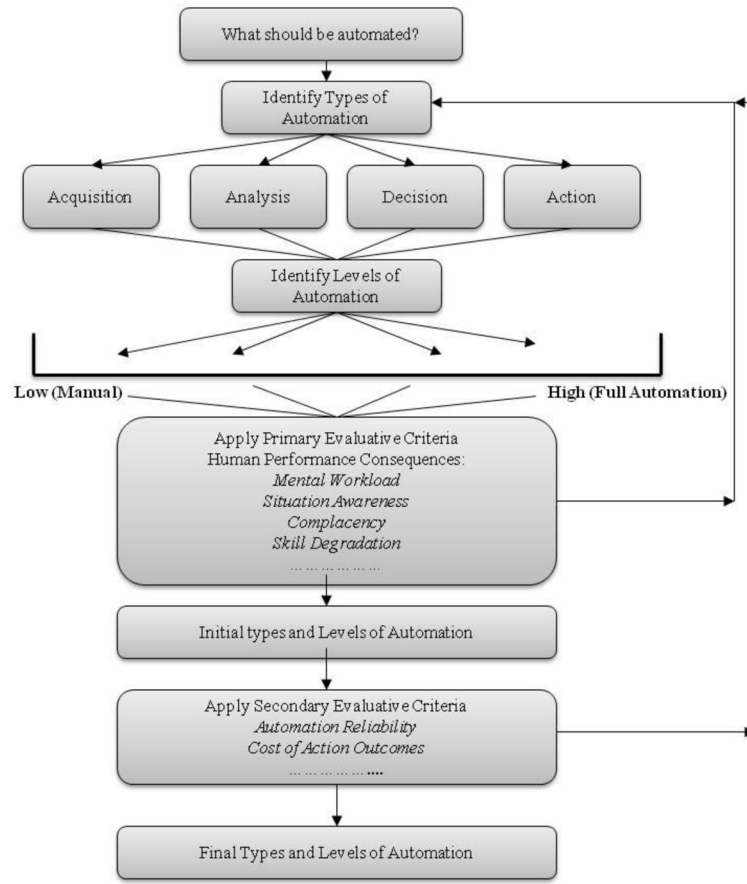


Figure 1. Flow chart showing application of the model of types and levels of automation (Parasurman, Sheridan, & Wickens, 2000).

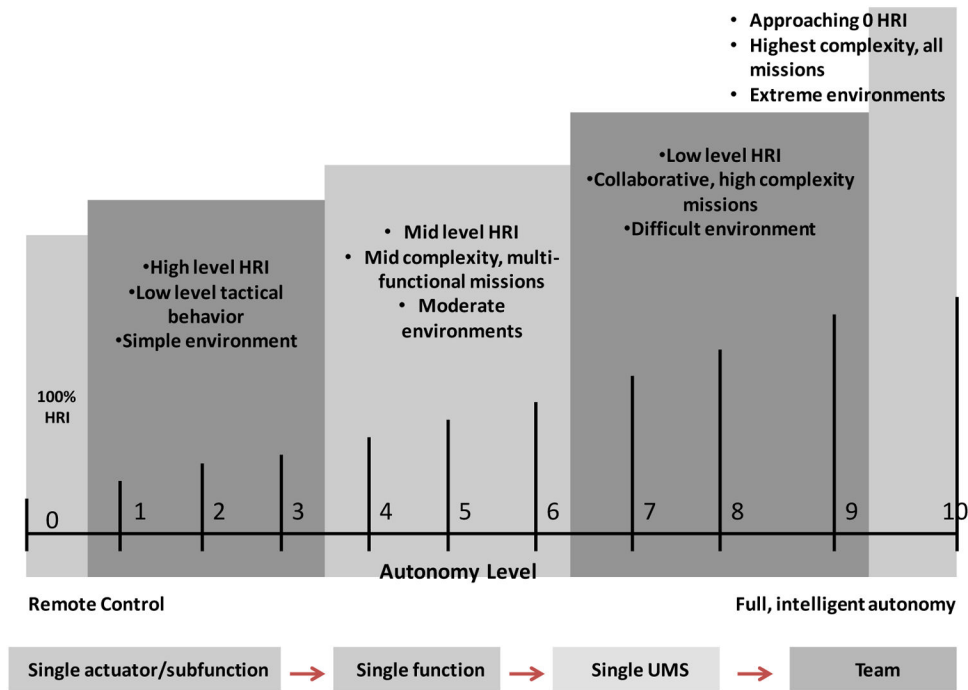


Figure 2. Autonomy Levels For Unmanned Systems (ALFUS) model of autonomy, depicting level of HRI along autonomy continuum (Huang, Pavek, Albus, & Messina, 2005).

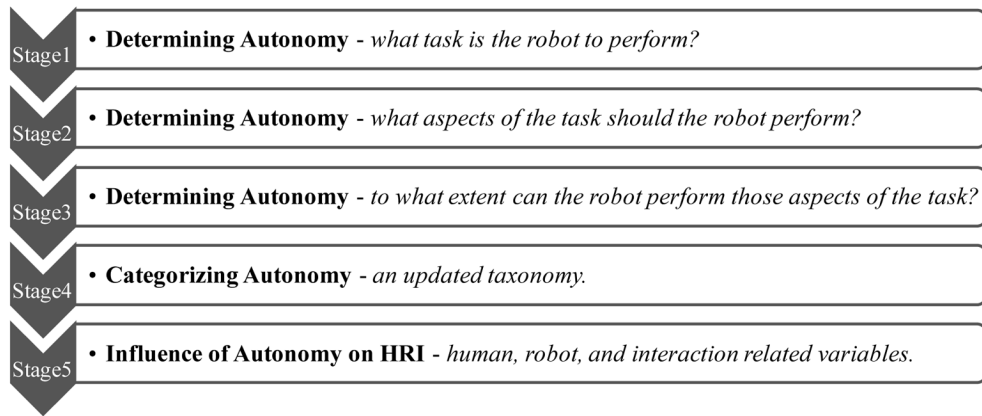


Figure 3.
Organizing flow chart to determine robot autonomy and effects on HRI.

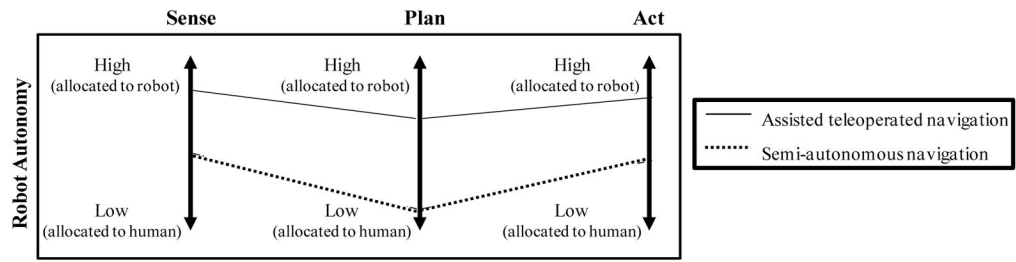


Figure 4. Levels of autonomy across the robot primitives sense, plan, and act. Two examples are given: assisted teleoperation (dotted line) and semi-autonomous navigation (solid line). Model modified from Parasuraman, Sheridan, and Wickens, 2000.

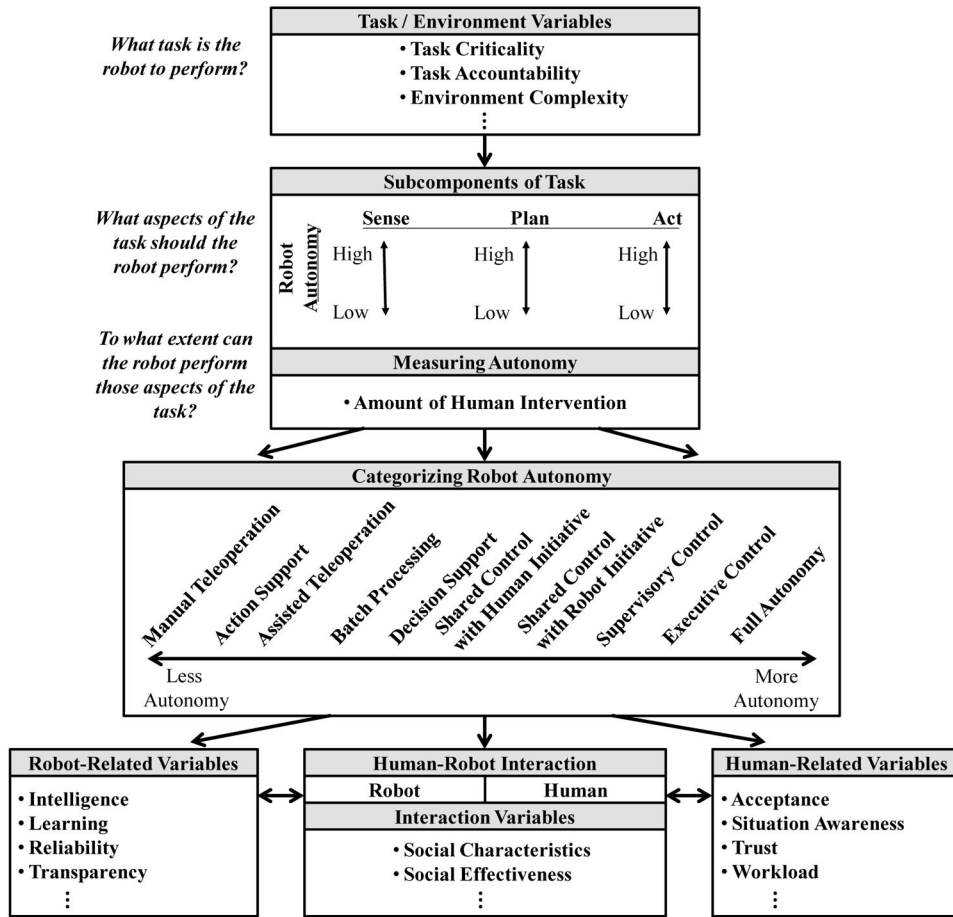


Figure 5. A framework of levels of robot autonomy for HRI. This framework can serve as a flow chart suggesting task and environmental influences on robot autonomy, guidelines for determining/measuring autonomy, a taxonomy for categorizing autonomy, and finally HRI variables that may be influenced by robot autonomy.

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, 1997, 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 |

Note: Emphasis (bold) added.

Table 2

Sheridan and Verplank (1978) Levels of Decision Making Automation

| Level of Automation | Allocation of Function |
|----------------------------|---|
| 1. | The computer offers no assistance; the human must take all decisions and actions. |
| 2. | The computer offers no assistance; the human must take 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 |

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Endsley and Kaber (1999) Levels of Automation

| Level of Automation | Description |
|-------------------------------|--|
| 1. Manual Control: | The human monitors, generates options, selects options (makes decisions) and physically carries out options. |
| 2. Action Support: | The automation assists human with execution of selected action. The human does perform some control actions. |
| 3. 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). |
| 4. 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. |
| 5. Decision Support: | The automation generates decision options that the human can select. Once an option is selected the automation implements it. |
| 6. Blended Decision Making: | The automation generates an option, selects it and executes it if they human consents. The human may approve of the option selected by the automation, select another or generate another option. |
| 7. 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. |
| 8. Automated Decision Making: | The automation selects and carries out an option. The human can have input in the alternatives generated by the automation. |
| 9. 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). |
| 10. Full Automation: | The system carries out all actions. |

Table 4

Proposed Taxonomy of Levels of Robot Autonomy for HRI

| Level of Robot Autonomy (LORA) | Function Allocation | | | Description | Examples from Literature |
|--|---------------------|-------|-----|---|--|
| | Sense | Plan | Act | | |
| 1. Manual Teleoperation | H | H | H | The human performs all aspects of task including sensing the environment and monitoring the system, generating plans/options/goals, and implementation. | “Manual Teleoperation” Milgram, 1995 “Tele Mode” Baker & Yanco, 2004; Bruemmer et al., 2005; Desai & Yanco, 2005 |
| 2. Action Support | 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” Kaber et al., 2000 |
| 3. Assisted Teleoperation | H/R | H | H/R | The human assist 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 |
| 4. Batch Processing | H/R | H | R | Both the human and robot monitor/sense the environment. The human, however, determines the goals and plans of the task. The robot then implements task. | “Batch Processing” Kaber et al., 2000 |
| 5. 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 robot to implement action. | “Decision Support” Kaber et al., 2000 |
| 6. Shared Control with Human Initiative | H/R | H/R | R | The robot autonomously senses the environment, develops plans/goals, and implements actions. However, the human monitors the robot’s progress, and may intervene and influence the robot with new goals/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” Tam et al., 1995 |
| 7. Shared Control with Robot Initiative | H/R | H/R | R | 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/plans. | “System-Initiative” Sellner et al., 2006 “Fixed-Subtask Mixed-Initiative” Hearst, 1999 |
| 8. Supervisory Control | H/R | R | R | Robot performs all aspects of task, but the human continuously monitors the robot. The human has over-ride capability and may set a new goal/plan. In this case the autonomy would shift to shared control or decision support. | “Supervisory Control” Kaber et al., 2000 |
| 9. Executive Control | R | (H)/R | R | The human may give an abstract high level goal (e.g., navigate in environment to specified location). The robot autonomously senses environment, sets plan, and implements action. | “Seamless Autonomy” Few et al., 2008 “Autonomous mode” Baker & Yanco, 2004; Bruemmer et al., 2005; Desai & Yanco, 2005 |
| 10. Full Autonomy | R | R | R | Robot performs all aspects of a task autonomously without human intervening with sensing, planning, or implementing action. | |

* Note: H = Human, R = Robot