A Taxonomy of Robot Autonomy for Human-Robot Interaction

Stephanie Kim
The University of Chicago
Chicago, USA
stephaniekim@uchicago.edu

Jacy Reese Anthis
The University of Chicago
Chicago, USA
anthis@uchicago.edu

Sarah Sebo
The University of Chicago
Chicago, USA
sarahsebo@uchicago.edu

ABSTRACT

Robot autonomy is an influential and ubiquitous factor in human-robot interaction (HRI), but it is rarely discussed beyond a one-dimensional measure of the degree to which a robot operates without human intervention. As robots become more sophisticated, this simple view of autonomy could be expanded to capture the variety of autonomous behaviors robots can exhibit and to match the rich literature on human autonomy in philosophy, psychology, and other fields. In this paper, we conduct a systematic literature review of robot autonomy in HRI and integrate this with the broader literature into a taxonomy of six distinct forms of autonomy: those based on robot and human involvement at runtime (*operational autonomy*, *intentional autonomy*, *shared autonomy*), human involvement before runtime (*non-deterministic autonomy*), and expressions of autonomy at runtime (*cognitive autonomy*, *physical autonomy*). We discuss future considerations for autonomy in HRI that emerge from this study, including moral consequences, the idealization of "full" robot autonomy, and connections to agency and free will.

CCS CONCEPTS

• **Human-centered computing** → **HCI theory, concepts and models**; • **General and reference** → *Surveys and overviews.*

KEYWORDS

Robot Autonomy, Human-Robot Interaction, Taxonomy, Robotics, Philosophy, Literature Review

ACM Reference Format:
Stephanie Kim, Jacy Reese Anthis, and Sarah Sebo. 2024. A Taxonomy of Robot Autonomy for Human-Robot Interaction. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24), March 11–14, 2024, Boulder, CO, USA.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3610977.3634993

1 INTRODUCTION

Robot autonomy strongly and unavoidably shapes user experiences in human-robot interaction (HRI). For example, users who fully control a robot's movements via teleoperation (e.g., drones, telepresence robots) tend to view the robot as merely a tool rather than a "colleague" [157]. On the other hand, users who interact with robots that operate independently (e.g., robot vacuums, robots that give directions in airports) tend to view them as distinct agents with their own intentions and goals [191]. Users who interact with robots that exhibit higher levels of cognitive autonomy (e.g., moral judgement [49], cheating [104], achieving individual goals [98]) readily anthropomorphize and thus respond to them in ways that they might respond to humans [52, 122].

Despite the importance of robot autonomy, there are currently two key challenges with the conceptualization and usage of the term "autonomy" in HRI. First, it is conceptually ambiguous. Many papers in HRI use autonomy only as a descriptor that applies to all robots in a general paradigm (e.g., "autonomous vehicle") [96, 105, 186], while other papers describe robot autonomy as a complex feature intertwined with notions of mind [188], agency [84], or free will [46]. These conceptual differences make certain empirical questions intractable because different conceptualizations yield different answers. For example, the finding that people are hesitant to enter an interaction with a robot may be explained by a reduction in user control if autonomy is conceptualized as a lack thereof [132] or explained by a threatening level of robot intelligence if autonomy is conceptualized as a sophisticated mental faculty [160].

Second, over the past two decades in HRI, when autonomy has been made conceptually precise, it has predominantly been a one-dimensional scale of the level of human involvement in robot operation, as in seminal taxonomies from Beer et al. [15] and Huang et al. [85]. This narrow view of robot autonomy stands in contrast to the literature outside HRI with rich and complex conceptualizations of human autonomy in philosophy, law, medicine, social science, and even in human-computer interaction (HCI), such as autonomy as a basic human right [66] and autonomy as the essence of what it means to be a person [59]. In HRI, we seem to be conflating, rather than articulating, critically different cases of autonomy. Consider that one study shows that increased robot autonomy leads to greater trust in the robot [154] while another describes how increased robot autonomy leads to less trust [10]. How can an increase in autonomy lead to opposite outcomes? We propose that both conclusions can be true simultaneously because the two works are referring to different forms of autonomy altogether. While the former appears to be measuring autonomy as the level of human intervention, the latter seems to be measuring it in terms of the robot's cognitive capabilities. If we can tease apart these cases, we will have a powerful conceptual tool to understand and predict user experience outcomes (e.g., user situational awareness [146]) and robot ethics [47].

Our approach addresses three research questions:

RQ1: How have other fields (e.g., philosophy, psychology) conceptualized and used the term "autonomy"?

RQ2: How has the field of HRI conceptualized and used the term "autonomy"?

HRI '24, March 11–14, 2024, Boulder, CO, USA

ACM ISBN 979-8-4007-0322-5/24/03.
https://doi.org/10.1145/3610977.3634993

RQ3: How can we develop a taxonomy that integrates these literatures to allow the HRI community to more precisely understand and communicate distinct forms of robot autonomy?

To answer these research questions, we first review popular conceptualizations of autonomy from outside the HRI literature and present a systematic literature review of the usage of "autonomy" in HRI. We then propose a comprehensive taxonomy based on these literatures with six distinct forms of autonomy summarized in Table 1: those based on robot and human involvement at runtime (*operational autonomy*, *intentional autonomy*, *shared autonomy*), human involvement before runtime (*non-deterministic autonomy*), and expressions of autonomy at runtime (*cognitive autonomy*, *physical autonomy*). Finally, we discuss how our taxonomy could allow the HRI community to better manage, predict, and revise expectations for human interactions with autonomous robots.

2 CONCEPTUALIZATIONS OF HUMAN AUTONOMY

The predominant definition of autonomy in HRI has been the degree to which a robot operates without continuous supervision or intervention from a human operator [7, 170]. High robot autonomy can thus be defined as low levels of human operator involvement (i.e., high levels of human operator disinvolvement), and low robot autonomy as high levels of human operator involvement [85]. This relatively narrow conceptualization in HRI is a sharp contrast with the varied and rich conceptualizations of human autonomy outside of robotics and HRI. Etymologically, "autonomy" is being subject to one's own laws, derived from the Greek *autos*, "self," and *nomos*, "law" or "governance." Scholars have interpreted this in myriad ways. We answer RQ1 with a brief overview, developed through the first and second authors aggregating various definitions of autonomy by and across disciplines into conceptual foci.

2.1 Self-Determination

Self-determination can refer to one's actions being caused by oneself rather than external forces [106]. Friedrich Nietzsche viewed a person as autonomous if and only if they are *causa sui*, or "self-causing" [61]. However, this arguably means no agent is truly autonomous, in the sense that all events have past causes [61].

Most other accounts are substantially weaker, only requiring, for example, that one is able to have some control over whether to act on certain desires. In the "hierarchical" model of autonomy [50], the self is identified with an internal freedom to endorse or reject first-order desires, which are desires for anything other than a desire, such as "I want to not overeat," but not higher-order desires, such as, "I want to not want to overeat" [50, 59]. To have "reflective self-evaluation" is to be aware of first-order desires and able to choose which to act upon based on higher-order desires [59, 71, 177]. The closely related "substantive" model of autonomy states that an absolute endorsement of our actions is not essential to autonomy. Instead, autonomy entails the capacity to revise our decisions with moral reasoning [50, 184]. Endorsement of actions need not be entirely self-originating in order to constitute self-determination, as the hierarchical model suggests, but one must be able to challenge the external forces that would otherwise determine one's own actions.

Self-determination is also central to psychological definitions of autonomy that stress the importance of a sense of self [44, 159]. Self-determination theory (SDT) is one of the most popular psychological frameworks for human motivation that comprises three basic needs: autonomy, competence (i.e., the ability to implement autonomy oneself), and relatedness (i.e., the ability to implement autonomy with others). The central distinction in SDT is between "autonomous" and "controlled" motivation. Autonomy is characterized in SDT as "volition, or a self-endorsement of their actions" [45]. Autonomous motivation originates from an internal locus of causality, which entails self-endorsement and the ability to choose one's decisions. The person's behavior is thus propelled by the identification of their actions with intrinsic values. Controlled motivation, by contrast, is dictated by external events and particularly by avoiding punishment from other people [149].

2.2 Independence

Many conceptualizations of autonomy focus instead on the direct independence of an individual's behavior from external forces, rather than the internalized sense of self and identity, though independence from external forces plays a role in that as well. These conceptualizations directly argue that the hallmark of a truly autonomous agent is to have liberty of judgment and choice from external dependencies—not just physical, but cognitive as well [40]. Philosophers of antiquity took this view. For example, Aristotle argued the self-sufficient person was not dependent on any other being or external force to guide their own happiness [50].

Legal and political conceptualizations also tend to emphasize independence from external forces, such as oppressive governments [9, 41]. HCI researchers primarily conceive of human autonomy as the human's degree of independence from the computer's restrictions [16]. Medical definitions emphasize the lack of external "hindrances" to the patient's cognition and physical activity [63].

2.3 Other Conceptualizations

In political theory, autonomy is also used to describe the self-governance of collective entities. Autonomy can be an attribute of political states, ethnic groups, territories, and institutions if they are able to operate without the control of a higher level of government, and is typically protected by international law [56]. Political autonomy can also entail the autonomy of individuals within the collective body, particularly the freedom for individuals to self-govern as long as it does not infringe on the autonomy of others. Philosopher Jean-Jacques Rousseau famously conceptualized the "general will," the interests of a collective that are derived from the individuals therein [17]. Thompson [165] argued that this is a form of autonomy. Through liberating oneself from individualistic, possibly hedonistic, motives and desires (i.e., the "private will") and assimilating into the general will, one preserves her individual autonomy [165]. It is only through the social and cultural environments to which one chooses to respond that an individual is able to critically evaluate her own values, beliefs, and desires. Under this view, individual autonomy cannot be conceived through self-evaluation alone [34].

Autonomy is frequently associated with "freedom of the will" or free will—which is distinct from Rousseau's idea of the general

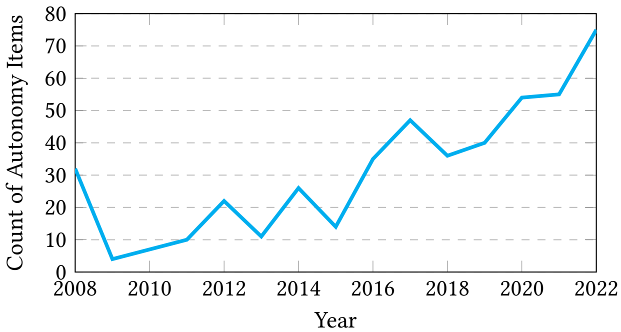

Figure 1: The total count of items from HRI, RO-MAN, IJSR, and THRI with "auton*" in the title, keywords, or abstract.

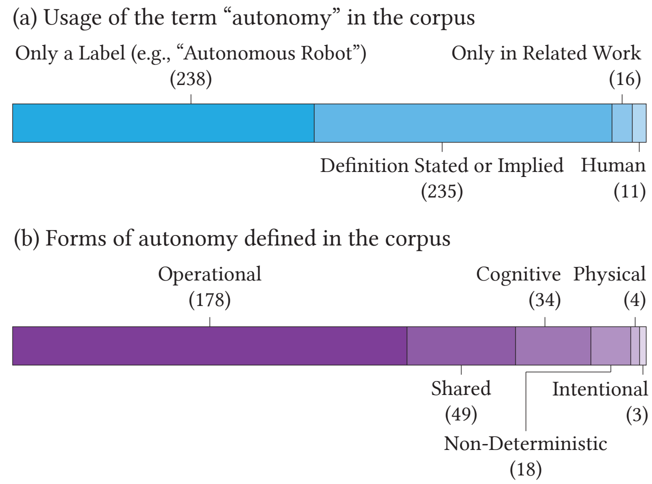

Figure 2: Usage and forms of "autonomy" in the reviewed corpus.

will. Free will is generally regarded as a degree of power or control over one's actions [125]. The nature of the relationship between free will and autonomy is rigorously debated. In some theories, free will is stronger than autonomy: an agent can be autonomous without having free will but cannot have free will without being autonomous [28]. In other theories, autonomy is the capacity to reflect upon and execute free will, providing the "self" that enables self-evaluation and self-governance. Kant is known for this view, arguing broadly that an autonomous will is a free one [90]. In Kantian ethics, autonomy is "made possible" by exerting free will to follow the universal moral law, or the "categorical imperative", and not being subject to any external moral laws or one's own first-order desires or inclinations [69, 91]. To be autonomous is to have a certain moral independence and freedom to fulfill one's duty [40, 91].

2.4 Takeaways from the Conceptualizations of Human Autonomy

We divide the usage of autonomy into two conceptual foci: self-determination and independence. Self-determination is a sense of direct internal control over one's actions, choices, and identity. Independence is the freedom from external control over one's cognitive and physical actions. Additionally, political autonomy is a way that a body of individuals can be collectively autonomous. The relationship between free will and autonomy raises discussion on what kinds and degrees of cognitive and moral capacities qualify one as autonomous. As we will later show, our taxonomy can leverage this literature on human autonomy to both broaden the conceptualization of robot autonomy in HRI and to make it more precise.

3 SYSTEMATIC LITERATURE REVIEW

To answer RQ2, we examined how robot autonomy is used within the HRI community. We conducted a systematic literature review of research articles published in four high-impact HRI venues: the ACM/IEEE International Conference on Human-Robot Interaction (HRI), the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), the International Journal of Social Robotics (IJSR), and ACM Transactions on Human-Robot Interaction (THRI). We reviewed all items with "auton*" (e.g., autonomy, autonomous) in the title, abstract, or keywords through July 31st, 2023. Because our focus is full research articles, we excluded the 262 HRI publications in other formats (e.g., extended abstracts, videos, late-breaking reports), resulting in 500 included papers.

A subset of 40 papers (10 from each venue) was annotated by four researchers to reach consensus on procedure and categorization. Once consensus was reached, the other 460 papers were each annotated by one researcher except when a categorization was ambiguous and needed input from others. For each paper that either defined autonomy (e.g., "the robot acted autonomously (without intervention from a human operator or Wizard of Oz, WoZ)") or implied a definition (e.g., "the robot must not require any human operators and must operate fully autonomously"), it was categorized into one or more forms of autonomy in HRI (e.g., operational, cognitive), which we developed over the course of the review process and describe in detail in Section 4. Each other paper was placed in one of three categories based on whether the term "auton*" was only used as a label for a robot (e.g., "autonomous social robot"), a summary of related work, or a reference to human rather than robot autonomy. Additionally, if any of the 500 papers included a human-subjects experiment, we annotated how autonomy was explicitly used as an independent variable or dependent variable. Finally, we took note of any papers that had a notable discussion of "agency" or derivative terms. The complete literature review annotations are available as a supplemental document.

As shown in Figure 1, there was a notable increase in the number of papers mentioning autonomy from 32 in 2008 to 75 in 2022. The year 2023 (30 papers) was not included in the graph because there was not a full year of data available. As shown in Figure 2(a), we found 235 explicit or implied definitions of autonomy (98 explicit, 137 implied). Another 238 papers only had robot autonomy as a label (e.g., "autonomous robot," "autonomous vehicle"); 16 only had robot autonomy as a summary of related work; and 11 referred only to human autonomy (whether defined, used as a label, or only referenced in a summary of related work). As displayed in Figure

2(b), of the different forms of robot autonomy, we categorized 178 as operational, 49 as shared, 34 as cognitive, 18 as non-deterministic, four as physical, and three as intentional. Of the 238 papers that only had autonomy as a label, which were most common in RO-MAN (60% of papers), we did not place them in one of these categories because they did not explicitly state or imply a definition, but we assumed that the papers' authors had the operational definition, if any, in mind. For example, when authors only used the label "autonomous vehicles," they seemed to have the definition in mind of a vehicle that travels without a human operator [20, 30, 92].

4 TAXONOMY OF ROBOT AUTONOMY

To answer RQ3, we propose a taxonomy that characterizes distinct forms of robot autonomy (see Table 1) by integrating the rich, interdisciplinary discussions of human autonomy with the results of our systematic literature review. We conceptualize three domains of robot autonomy. Human and robot involvement at runtime includes *operational*, *shared*, and *intentional* autonomy; human involvement before runtime includes *non-deterministic* autonomy; and expressions of robot autonomy at runtime include *cognitive* and *physical* autonomy.

4.1 Human and Robot Involvement at Runtime

The first domain of robot autonomy in the taxonomy is the degree of human and robot involvement at runtime, which includes three forms of autonomy: *operational*, *intentional*, and *shared autonomy* (see Figure 3). This domain is *general-sum*, incorporating both human and robot involvement as independent factors. This perspective stands in contrast to the *zero-sum* perspective in many HRI models, which assumes, for example, that 90% human involvement implies 10% robot involvement [85, 123]. Some recent work in HCI has found a general-sum approach more useful for characterizing human-AI interactions, defining AI roles based on the levels of both "human involvement" and "AI autonomy" [93].

This conceptualization is displayed in Figure 3. We depict operational autonomy as the degree of human operator disinvolvement at runtime (the horizontal axis) and intentional autonomy as the degree of robot goal-oriented involvement at runtime (the vertical axis). This allows us to categorize common robots in terms of their location in the space. Teleoperated robots have low robot involvement (i.e., low intentional autonomy) and low human operator disinvolvement (i.e., high human involvement; low operational autonomy) [132]; they are essentially vehicles for their human operators and would be idle without one. On the other hand, a drone swarm typically has high robot involvement and low human operator disinvolvement [95]. The ability to differentiate cases such as teleoperated robots and drone swarms—each of which has low human disinvolvement, but different levels of robot involvement—is the key benefit of our general-sum taxonomy.

Shared autonomy is represented by the grey square, requiring at least some intentional involvement by both the robot and human. If either the robot or human has no control over operation (low intentional or high operational autonomy, respectively), there is no shared autonomy. Shared control paradigms, such as "symbiotic autonomy", where humans and robots "fill in" the other's mutual weaknesses [176], and negotiation, where both the human and robot "negotiate" their level of autonomy given the other's [89], are examples of shared autonomy.

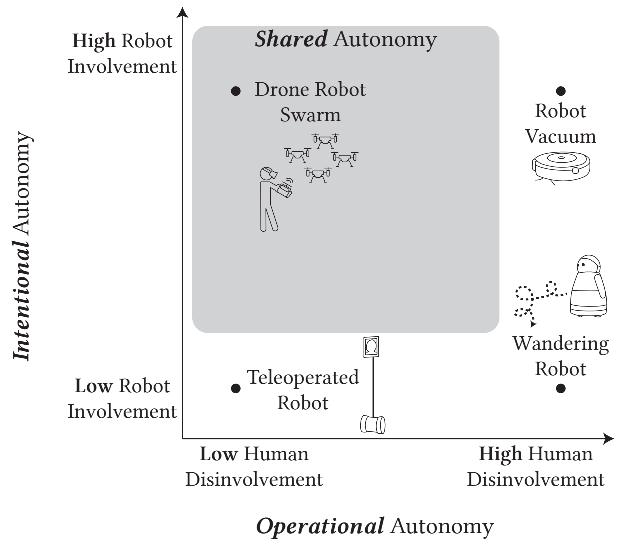

Figure 3: Robot autonomy at runtime is based on its degrees of human operator disinvolvement (*operational autonomy*) and robot goal-oriented involvement (*intentional autonomy*). Intentional involvement from both human and robot at runtime represents *shared autonomy*, depicted as the grey rectangular area.

4.1.1 *Operational autonomy.*

We define operational autonomy as the degree of human operator disinvolvement at the robot's runtime. We refer to "disinvolvement" instead of "involvement" for consistency with the other forms, in which "high" denotes high autonomy and "low" denotes low autonomy. Full operational autonomy occurs when there is full human operator disinvolvement (i.e. no human operator involvement), so the robot is able to function and execute its tasks entirely on its own. For example, a Roomba robot vacuum can clean an assigned area without any human intervention. On the other hand, in the Wizard of Oz (WoZ) paradigm or with telepresence robots, there is little to no operational autonomy because all robot actions are conducted by a human operator. Olatunji et al. [131] say that autonomy is "typically considered as a continuous or discrete spectrum, with direct human control and full autonomy at either end, and any number of intermediate levels in between." Our conceptualization of operational autonomy is inspired by the aforementioned conceptual focus of autonomy as "independence" from external factors. Relative to the robot, human intervention is external, even if it is in response to the robot's needs within its local environment (e.g., a broken wheel). Even when operationally autonomous, robots are usually inexorably tied to the context of their environment [24] and human operators [82].

Operational autonomy is the dominant conceptualization of autonomy in the HRI literature. Figure 2 shows that 178 out of 235, or 75.7%, of papers that define autonomy do so as operational. Explicit definitions focus on the omission or inclusion of human involvement, using phrases such as "amount of/without

Term	Definition	Example	Associated Terms
Human and Robot Involvement at Runtime			
Operational Autonomy	the degree of human operator disinvolvement at runtime	• "no human intervention" [42] • "do its tasks without a human's input" [53] • "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" [15]	control, input, intervention, involvement, remote, tele-operation, tele-presence, Wizard-of-Oz
Intentional Autonomy	the degree of robot goal-oriented involvement at runtime	• "several types of in-the-wild studies of mobile robots can be run without autonomous navigation, using wandering instead" [120]	deliberate, directional, goal, random, wandering
Shared Autonomy	the degree to which the human operator and robot are both intentionally involved at runtime	• "joint control of a robot by a human user and an autonomous control system" [26] • "reduce the amount of control people need to exert by 1) predicting people's goals and 2) taking assistive actions toward those goals" [7]	adjustable, degrees, levels, semi-autonomous, shared control, symbiotic
Human Involvement Before Runtime			
Non-deterministic Autonomy	the degree to which a robot's behavior is not specified prior to runtime	• "autonomous or pre-programmed" [84] • "neither remote controlled by a human, nor scripted previously, but is instead generated by an algorithm that reacts to external sensor stimuli" [151]	adaptive, determine, pre-programmed, predict, scripted, stimuli
Expressions of Autonomy at Runtime			
Cognitive Autonomy	the degree to which the robot takes cognitive action	• "to make ethical choices" [49] • "to reason, or understand the sensory information it is receiving, and plan the most appropriate action based on this input" [189]	choice, decision, goal, moral, reason, think
Physical Autonomy	the degree to which the robot takes physical action	• "e.g., [...] is the robot confined to a pen or does it wander and get stuck under television cabinets" [150] • "energetic autonomy, i.e., the capability to cover long distances without the need to be recharged" [115]	confined, dance, drive, move, stuck

Table 1: We propose a comprehensive taxonomy of six distinct forms of robot autonomy based on our systematic literature review and integration with the literatures on human autonomy in other fields (e.g., philosophy, psychology).

human/participant/operator intervention" [7, 39, 53, 58, 70, 170], "without help/assistance from a human/operator" [22, 166], "controlled/not controlled by a human" [60, 83, 117, 133], and "sensing its human partner" [136]. Operational autonomy is often implied by contrast to WoZ [23, 81, 88], teleoperation [119, 147, 173, 190], or "operator," "manual," or "remote" involvement or control [8, 32, 54, 67, 86, 121, 141, 180]. Operational autonomy is also frequently implied by referencing previous work that entails operational autonomy, such as the Levels of Robot Autonomy (LoRA) scale [15] or the ALFUS taxonomy [85], which state that if the human operator is at 100% involvement, the robot must be at 0%. Autonomous vehicles papers often reference the scale made by the Society of Automotive Engineers [80, 130, 164], which defines six levels of driving automation (i.e., level of driver disinvolvement) from no driving automation (Level 0) to full driving automation (Level 5). Similar scales exist in medical robotics [187] and unmanned systems research [85].

Operational autonomy is also the form of autonomy most frequently manipulated in HRI studies [48, 109, 163]. Manipulations tend to present discrete levels of robot autonomy as defined in the LoRA scale [15], either varying on whether humans are involved at runtime at all [170] or on a continuum of low, partial, or high input [4, 18, 48, 65, 109, 121, 123, 131, 163, 188]. A common set-up compares an autonomous robot with one that is WoZ or teleoperated [33, 81, 119, 121, 168]. Some papers describe a "semi-autonomous" approach, wherein a robot can recover from errors made in its autonomous state through operator intervention [94, 156]. Brooks et al. [27] varied the length of time in which the human operator "ignores" or "neglects" the robot with longer times constituting higher autonomy. Studies have found various effects of operational autonomy, such as Torre et al.'s [167] finding that participants are less likely to swerve out of the way to avoid collision with autonomous robots than teleoperated robots. Outcomes of interest include performance [32, 131] and perceptions of operational autonomy and robot reliability [48]. For instance, Rosenthal-von der Pütten et al. [143] found that more human-like robot appearance and the inclusion of a narrative about the robot, rather than just an instruction manual, led to evaluations of the robot as more autonomous and intelligent. Reuben et al. [147] tested participants' judgements of whether a robot was autonomous or teleoperated based on their mental model of the robot over a six-week in-the-wild study and found that participants' judgements were impacted by their understanding of possible behaviors of autonomous robots.

4.1.2 Intentional autonomy. We define intentional autonomy as the degree of robot goal-oriented involvement, or simply robot involvement, at runtime. Intentional autonomy is analogous to operational autonomy, but it is determined within the robot and without direct reference to external factors. Our conceptualization of intentional autonomy is thus associated with self-determination, because intentional autonomy is about the robot's internal, rather than externally imposed, goals. Because the robot is necessarily involved with its own behavior in the sense that it is the one exhibiting the behavior, we specify "goal-oriented involvement" as the robot being both oriented towards a particular goal and having knowledge of this goal. We see the robot knowing its goal, as opposed to accidentally achieving it, as having the necessary internal representation to stay targeted at the goal. For example, Nanavati et al. [120] describe a "wandering" robot that arbitrarily explores its local environment with no directionality. In this case, while there is full operational autonomy (i.e., full human operator disinvolvement), if the robot ascribes no purpose or internal representation to its "wandering", there is no intentional autonomy. In the absence of any involvement, behavior is left to random chance or environmental factors (e.g., wind, bystanders). Robots with high intentional autonomy exhibit non-random behavior at runtime [103]. For example, an effective game-playing robot that evaluates the game state and selects actions to optimize its chance of winning [2] would have high intentional autonomy.

In this taxonomy, intentional autonomy does not require specific mental faculties, such as self-awareness or agency, as "intention" is sometimes conceptualized outside of HRI [182], only that the robot knows its goal in the previously defined sense. One could argue that no robots achieve this because they are by definition created by humans and take action according to human inputs. However, it is clear in the human case that, when humans are willfully or forcibly constrained based on another person's intentions (e.g., a child grounded by their parents), they maintain some autonomy. To speak in terms of our conceptual foci, their independence may be limited, but their self-determining faculties (e.g., cognitive abilities, goal-setting) can still manifest. With intentional autonomy, the idea is that even though robots may not escape the influence of a human at any point in its operation, robots can still express purposeful behavior to achieve some goal. We can observe this when, for example, a robot self-determines the best "level of autonomy" to use [146].

4.1.3 Shared autonomy. We define shared autonomy as the degree to which the human operator and robot are both intentionally involved at runtime. Shared autonomy is unique in our taxonomy, in that it is a composite of operational and intentional autonomy, as shown in Figure 3. High shared autonomy occurs when the human and robot are more equally involved, and low shared autonomy is when one is much more involved than the other. This concept of shared autonomy corresponds to myriad terms in the HRI literature: symbiotic autonomy [12, 176, 179], shared-autonomy teleoperation [62], adjustable autonomy [64, 77, 108, 118], adaptive autonomy [11, 75], shared control [26, 58, 89, 127, 172], assistive robot control [7], traded control [126], semi-autonomy [13, 21, 38, 94, 99, 107, 114, 138, 139, 156, 158, 169, 175, 185], confidence-based autonomy [161], collective autonomy [35], and shared autonomy [7, 26, 72, 98, 116, 124]. Our conceptualization is partly influenced by the outside literature on collective autonomy. Some studies in HRI refer to collective autonomy in the context of "teams" or "swarms" of more than two systems collaborating [95, 110, 134], even explicitly circumscribing collective autonomy as being under the command of a single operator [35]. Outside of HRI, however, collective autonomy can encompass a unique sort of autonomy altogether for collections of individuals (e.g., the general public).

Shared autonomy manifests in two ways: *simultaneous control* and *exchanged control*. In *simultaneous control*, the human operator and the robot take action at the same time [94, 138, 185]. For instance, a robotic arm picking items off of a shelf may perform the movements themselves without human involvement, but with the direction and speed of the task being simultaneously dictated by the human operator. In *exchanged control*, only one party takes action at a time [6, 26, 89, 123]. For example, if a human picks up their robot vacuum and sets it in each room of a house, where the robot vacuum cleans without human involvement, this is a case of exchanged control between the human while moving between rooms and the robot while in a room. Most papers that define shared autonomy, or a closely associated concept, emphasize its benefits, particularly its flexibility in either the robot or human deciding when to take or yield control. Veloso [179] defines symbiotic autonomy as "the robots explicitly include actions to ask for help from humans in their behavior policies." Similarly, Naghsh et al. [118] define adjustable autonomy as one in which, "the robot entity needs not make all decisions autonomously, rather it can choose to reduce its own autonomy level and transfer decision making control to other users or agents." Supervision can also serve to correct errors [112], and many researchers describe the human-robot cooperation as a "blend" of user and robot involvement [26, 32, 57, 116, 127]. As robots become more sophisticated and collaborative (e.g., "Human-Autonomy Teams" [1]), questions of collective autonomy will become more pertinent–not just for robot autonomy, but also how robots should be deployed when their autonomous operation may affect the collective autonomy of political states and other institutions.

4.2 Human Involvement Before Runtime

During runtime (i.e., the period in which the robot is operational), both the human operator and robot can have input and control over the robot's autonomy. Prior to runtime, however, only the human

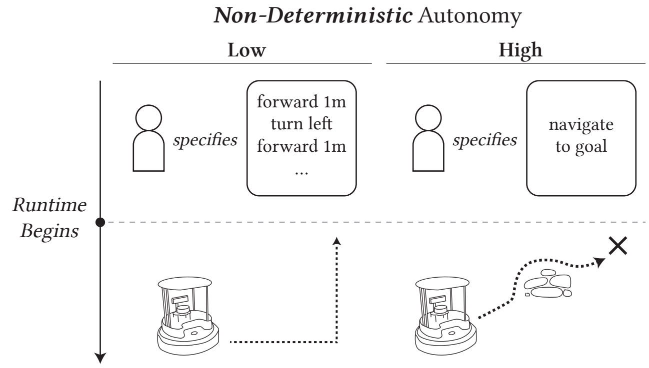

Figure 4: Non-deterministic autonomy is the degree to which a robot's behavior is not specified prior to runtime, as shown by the example of a navigational robot.

takes action since the robot is not operational. Therefore, there is only one axis of autonomy, which we characterize in this section as *non-deterministic autonomy*.

4.2.1 Non-deterministic autonomy. We define non-deterministic autonomy as the degree to which the robot's behavior is not specified prior to runtime. Figure 4 (left) illustrates a robot with low non-deterministic autonomy because it is given exact instructions to reach its destination; there is no leeway for the robot to further optimize or diverge from its navigation path. Examples of low non-deterministic autonomy in the literature include robots that are explicitly said to be "pre-programmed," [3, 84, 102] "scripted previously," [151] or have "predefined" [31] or "manually designed" [111] behavior. On the other hand, the robot with high non-deterministic autonomy in Figure 4 (right) is not provided the exact sequence and execution of its runtime operation. Instead, the robot can decide the optimal path during runtime to reach its destination and adapt to novel situations at runtime, such as "what information is needed and at what time and place it is requested" [73], in order to accomplish its goal. Examples of highly non-deterministic robots in the literature include those that are able to adapt to their environment [31, 78, 140, 162] or adapt to people [79, 142] in real-time.

In practice, non-deterministic autonomy may correlate strongly with intentional autonomy. However, we view these two forms of autonomy as distinct. For example, a robot that has extensive specification of behavior prior to runtime may behave in a goal-oriented way during runtime (e.g., navigating to a specific location despite alternative rewards), but that behavior has still been specified prior to runtime. The robot would thus have low non-deterministic autonomy but high intentional autonomy.

It remains unclear exactly how non-deterministic autonomy could be operationalized or measured. One could measure the degree of conditionality (e.g., number of "if" statements in the code), the diversity of possible runtime environments, or the sophistication of action selection subprocesses, but in each case the robot's programming would still determine its actions in a given context. Nonetheless, this concept seems to capture an important sense of autonomy in the HRI literature that varies even if the robot has full operational or intentional autonomy.

4.3 Expressions of Robot Autonomy at Runtime

In addition to considering robot and human involvement in a robot's operation before and during runtime, there are different ways in which robot autonomy is expressed at runtime, which we categorize as *cognitive* and *physical*.

4.3.1 Cognitive autonomy. We define cognitive autonomy as the degree to which the robot takes cognitive action. Generally, cognition is defined as "[t]he action or faculty of knowing taken in its widest sense, including sensation, perception, conception, etc., as distinguished from feeling and volition" [36]. Robot cognition includes the ability to make and execute decisions [135, 148, 155], set and achieve goals [15, 98], understand others' mental and affective states [14, 108, 128], act upon one's senses and planning [15, 37, 101], and learn from the environment and past experiences [19, 86, 144, 189]. Discussion of robot cognition is rife with the same sort of disagreement as has persisted for decades in the human cognition literature [97]. Some have explicitly defined robot cognition as synonymous with artificial intelligence and autonomy [77]. We view cognitive autonomy as not requiring intentional autonomy. It is possible for a robot to reason about the world in complex ways without a particular goal in mind.

In our literature review, forms of cognitive autonomy fell naturally into two categories: moral autonomy and decisional autonomy. *Moral autonomy* means the robot is capable of using moral reasoning to select actions, such as distinguishing right from wrong [25, 49, 129, 174] and resolving morally ambiguous scenarios [160]. Moral autonomy often involves an awareness of ethical and social norms prior to operation [68, 181] and of moral stakeholders during operation [133, 171]. *Decisional autonomy* is a robot's decision-making ability, such as the "ability to observe and act on their environment, as well as conduct activities toward achieving both individual and collective goals" [98] or to "generate its own strategies for action" [37] and independently evaluate or deliberate on those actions [55, 101, 135]. For example, decisionally autonomous robots can independently reason about "manipulation goals" when predefined goals by the human operator may fail [87]. A particularly important decision-making ability is the "ability to disregard human commands" [2], which was only utilized in one paper in our review [2], but several included papers cited Złotowski et al. [191], who utilized this form of decision-making.

4.3.2 Physical autonomy. We define physical autonomy as the degree to which the robot takes physical action. The paradigmatic case of restricting physical autonomy is confining the robot in a physical space, such as a cage [150], which we refer to as *movement autonomy*. Levillain et al. [100] discuss autonomy as "bursting into sudden movements," which would require significant freedom of movement. Physical autonomy can also manifest through the resource requirements of the robot, particularly its energy source, which we refer to as *energetic autonomy*. An example of a robot with high energetic autonomy is a drone that can recharge itself to fly indefinitely [115]. Physical autonomy can be increased or decreased based on the desired robot behavior. In a study by Horn et al. [83], the same robot could be used either in a "passive mode" as a walking aid or in an an "autonomous mode" wherein the robot can move towards the user without human assistance.

5 FUTURE CONSIDERATIONS FOR ROBOT AUTONOMY

In this section, we discuss future considerations for autonomy in HRI that emerged from this work: moral consequences, the idealization of "full" robot autonomy, and connections to agency and free will. Developing these perspectives will allow the HRI community to better manage, predict, and revise expectations for the scaling and enrichment of human interactions with autonomous robots.

5.1 Moral Consequences of Robot Actions

We defined moral autonomy as a robot's capability to select actions through moral reasoning, but there is a wide spectrum between merely having behavior with moral consequence, such as being pre-programmed to encourage a human patient to take their medication [174], and human-level consideration of right and wrong, such as in Kantian moral theory through adherence to universal moral law. In both cases, HRI designers face difficult decisions in choosing between ethical frameworks like utilitarianism and deontology, and these challenges will only increase as increased autonomy involves robots making more decisions with significant moral implications [49, 181]. An initial step towards tackling these challenges is addressing how robot autonomy can both erode and enhance the autonomy of humans. Robots shifting roles from "tool" to "colleague" [157] and gaining humans' trust [153] and moral consideration [5, 76, 137] may make humans feel as though their autonomy is threatened by and traded off with robot autonomy [18], or it could complement and preserve human autonomy by providing aid [183] and maximizing the user's capabilities, especially on command [43, 178].

5.2 Moving Towards "Full Autonomy"

Within their study, Wright et al. [186] state "full autonomy is an essential objective," and it is generally regarded as desirable but "not yet practical" [29] or "far from reality" [62]. Robot autonomy will scale towards this "fullness" in different ways and towards different forms. Operational autonomy is relatively straightforward: full operational autonomy refers to states where the degree of human operator disinvolvement is near 100% [51]. However, "full" autonomy is unclear for the other five forms we describe. In general, HRI researchers have expressed that robots are not, and should not be, able to possess free will [46] or commit harm against humans [113]. And yet, some argue that, "as the cognitive, perceptual, and motor capabilities of robots expand, they will be expected to be explicit ethical agents" [174]. As HRI research strives to keep pace with the real-world increase in robot autonomy, the ideals of human autonomy can be a useful yardstick. For instance, just as some accounts of human autonomy are based in reflective self-evaluation, more "advanced" robot autonomy, and its regulation, could be based in the robot's ability to reflect upon and criticize its choices.

5.3 Robot Agency and Free Will

We focus on autonomy because of its importance in HRI, its relative concreteness, and the lack of existing comprehensive frameworks. However, there are related concepts that also bear on the future of HRI. Agency is arguably the most closely related concept, as reflected in the numerous discussions of agency in the reviewed items. For example, Zafari and Koeszegi [188] define agency as "the capacity to perform a goal-oriented task to an extent autonomously on the environment," implying that agency requires autonomy, but others define autonomy and agency as separate concepts [132, 145]. In many papers focused on reinforcement learning, the word "agent" is used merely to differentiate the subject of the model from its environment, which presumably does not require autonomy in any of the definitions considered in this review. This minimal sense of agency is also the most common outside of HRI, in which an agent is merely a system that takes action [152]. Bennett et al. [16] provide a thorough and timely review of agency and autonomy in the context of HCI, though their focus is on that of the human, rather than the computational or robotic system.

We did not often find the concept of "free will" in the reviewed items. The outside literature has many different conceptualizations of free will, and there seems to be agreement that it is a more sophisticated and less common property compared to autonomy. This helps explain the lack of discussion in HRI and robotics, because while robots are increasingly complex and capable, few would argue that they are sophisticated enough to possess free will [46]. If we adapted the Kantian notion of autonomy to robots, such that autonomy relied on free will, then robot autonomy is not even a possibility for current systems. Human-like robots such as Sophia, the world's first robot with legal citizenship, make salient the possible futures of robots with roles, responsibilities, and even rights [74]. HRI needs rigorous frameworks of autonomy, agency, and free will to answer impending questions of moral responsibility, fairness, accountability, and transparency.

6 CONCLUSION

Robot autonomy influentially and ubiquitously shapes the user experience in HRI, yet its conceptualization and usage within HRI faces two primary challenges: conceptual ambiguity and unidimensionality. In this paper, we addressed these challenges by broadening the conceptualization of robot autonomy in HRI. We conducted a systematic literature review of robot autonomy in HRI and integrated this with interdisciplinary literatures outside of HRI to propose a taxonomy with six distinct forms of autonomy: those based on robot and human involvement at runtime (*operational autonomy*, *intentional autonomy*, *shared autonomy*), human involvement before runtime (*non-deterministic autonomy*), and expressions of autonomy at runtime (*cognitive autonomy*, *physical autonomy*). A number of future considerations for autonomy in HRI emerged from this process, including moral consequences, the idealization of "full" robot autonomy, and connections to agency and free will. Going forward, we encourage HRI researchers to clarify which sort of autonomy they are analyzing in their conceptual and empirical studies, as well as to extend and critique this taxonomy to better evaluate the evolving nature and effects of autonomous robots in human society.

ACKNOWLEDGMENTS

We thank Madeline Busse and Aurelie Roubinowitz for their help in annotating the papers contained in our systematic literature review.

REFERENCES

[1] Gene M. Alarcon, Joseph B. Lyons, Sarah A. Jessup, et al. 2023. Affective responses to trust violations in a human-autonomy teaming context: humans versus robots. *International Journal of Social Robotics*, 1–13.

[2] D. D. Allan, Andrew J. Vonasch, and Christoph Bartneck. 2022. Better than us: the role of implicit self-theories in determining perceived threat responses in hri. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 215–224. doi: 10.1109/HRI53351.2022.9889520.

[3] Fady Alnajjar, Massimiliano Cappuccio, Abdulrahman Renawi, Omar Mubin, and Chu Kiong Loo. 2021. Personalized robot interventions for autistic children: an automated methodology for attention assessment. *International Journal of Social Robotics*, 13, 67–82.

[4] Antonio Andriella, Henrique Siqueira, Di Fu, Sven Magg, Pablo Barros, Stefan Wermter, Carme Torras, and Guillem Alenyà. 2021. Do i have a personality? endowing care robots with context-dependent personality traits. *International Journal of Social Robotics*, 13, (Dec. 2021), 1–22. doi: 10.1007/s12369-020-0069 0-5.

[5] Jacy Reese Anthis and Eze Paez. 2021. Moral circle expansion: A promising strategy to impact the far future. *Futures*, 130, (June 2021), 102756. doi: 10.101 6/j.futures.2021.102756.

[6] Junki Aoki, Ryota Yamashina, and Ryo Kurazume. 2021. Teleoperation method by illusion of human intention and time. In *2021 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 482–487. doi: 10.1109/RO-MAN50785.2021.9515457.

[7] Reuben M. Aronson, Thiago Santini, Thomas C. Kübler, Enkelejda Kasneci, Siddhartha Srinivasa, and Henny Admoni. 2018. Eye-hand behavior in human-robot shared manipulation. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (HRI '18). Association for Computing Machinery, Chicago, IL, USA, 4–13. isbn: 9781450349536. doi: 10.1145/317122 1.3171287.

[8] Dante Arroyo and Fumihide Tanaka. 2018. A time-based strategy for the transition of control in telepresence robots. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 342–347. doi: 10.1109/ROMAN.2018.8525639.

[9] Shiri Azenkot, Sanjana Prasain, Alan Borning, Emily Fortuna, R.E. Ladner, and Jacob Wobbrock. 2011. Enhancing independence and safety for blind and deaf-blind public transit riders. In (May 2011), 3247–3256. doi: 10.1145/19789 42.1979424.

[10] Franziska Babel, Johannes Kraus, and Martin Baumann. 2022. Findings from a qualitative field study with an autonomous robot in public: exploration of user reactions and conflicts. *International Journal of Social Robotics*, 14, 7, 1625–1655.

[11] Gregory Bales and Zhaodan Kong. 2022. Neurophysiological and behavioral differences in human-multiagent tasks: an eeg network perspective. *J. Hum.-Robot Interact.*, 11, 4, Article 42, (Sept. 2022), 25 pages. doi: 10.1145/3527928.

[12] Kim Baraka and Manuela M Veloso. 2018. Mobile service robot state revealing through expressive lights: formalism, design, and evaluation. *International Journal of Social Robotics*, 10, 65–92.

[13] Juan Antonio Barragan, Daniela Chanci, Denny Yu, and Juan P. Wachs. 2021. Sachets: semi-autonomous cognitive hybrid emergency teleoperated suction. In *2021 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1243–1248. doi: 10.1109/RO-MAN50785.2021.951 5517.

[14] Andrea Bauer et al. 2009. The autonomous city explorer: towards natural human-robot interaction in urban environments. *I. J. Social Robotics*, 1, (Apr. 2009), 127–140. doi: 10.1007/s12369-009-0011-9.

[15] Jenay M. Beer, Arthur D. Fisk, and Wendy A. Rogers. 2014. Toward a framework for levels of robot autonomy in human-robot interaction. *J. Hum.-Robot Interact.*, 3, 2, (June 2014), 74–99. doi: 10.5898/JHRI.3.2.Beer.

[16] Dan Bennett, Oussama Metatla, Anne Roudaut, and Elisa D. Mekler. 2023. How does HCI understand human agency and autonomy? In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, (Apr. 2023). doi: 10.1145/3544548.3580651.

[17] Christopher Bertram. 2023. Jean Jacques Rousseau. In *The Stanford Encyclopedia of Philosophy*. (Summer 2023 ed.). Edward N. Zalta and Uri Nodelman, (Eds.) Metaphysics Research Lab, Stanford University.

[18] Tapomayukh Bhattacharjee, Ethan K. Gordon, Rosario Scalise, Maria E. Cabrera, Anat Caspi, Maya Cakmak, and Siddhartha S. Srinivasa. 2020. Is more autonomy always better? exploring preferences of users with mobility impairments in robot-assisted feeding. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (HRI '20). Association for Computing Machinery, Cambridge, United Kingdom, 181–190. isbn: 9781450367462. doi: 10.1145/3319502.3374818.

[19] Alexandru Blidaru, Stephen L. Smith, and Dana Kulić. 2018. Assessing user specifications for robot task planning. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 72–79. doi: 10.1109/ROMAN.2018.8525546.

[20] Cara Bloom and Josiah Emery. 2022. Privacy expectations for human-autonomous vehicle interactions. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1647–1654. doi: 10.1109/RO-MAN53752.2022.9900615.

[21] Jonathan Bohren, Chris Paxton, Ryan Howarth, Gregory D. Hager, and Louis L. Whitcomb. 2016. Semi-autonomous telerobotic assembly over high-latency networks. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 149–156. doi: 10.1109/HRI.2016.7451746.

[22] Annika Boos, Markus Zimmermann, Monika Zych, and Klaus Bengler. 2022. Polite and unambiguous requests facilitate willingness to help an autonomous delivery robot and favourable social attributions. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1620–1626. doi: 10.1109/RO-MAN53752.2022.9900870.

[23] Cynthia Breazeal, Nick DePalma, Jeff Orkin, Sonia Chernova, and Malte Jung. 2013. Crowdsourcing human-robot interaction: new methods and system evaluation in a public environment. *J. Hum.-Robot Interact.*, 2, 1, (Feb. 2013), 82–111. doi: 10.5898/JHRI.2.1.Breazeal.

[24] Yohan Breux, Sebastien Druon, and Rene Zapata. 2018. From perception to semantics: an environment representation model based on human-robot interactions. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 672–677. doi: 10.1109/ROMAN.2018.8 525527.

[25] Gordon Briggs and Matthias Scheutz. 2014. How robots can affect human behavior: investigating the effects of robotic displays of protest and distress. *International Journal of Social Robotics*, 6, 343–355.

[26] Connor Brooks and Daniel Szafir. 2019. Balanced information gathering and goal-oriented actions in shared autonomy. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 85–94. doi: 10.1109/HRI.2019.8 673192.

[27] Daniel J. Brooks, Momotaz Begum, and Holly A. Yanco. 2016. Analysis of reactions towards failures and recovery strategies for autonomous robots. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 487–492. doi: 10.1109/ROMAN.2016.7745162.

[28] Sarah Buss and Andrea Westlund. 2018. Personal Autonomy. In *The Stanford Encyclopedia of Philosophy*. (Spring 2018 ed.). Edward N. Zalta, (Ed.) Metaphysics Research Lab, Stanford University.

[29] Maria E. Cabrera, Tapomayukh Bhattacharjee, Kavi Dey, and Maya Cakmak. 2021. An exploration of accessible remote tele-operation for assistive mobile manipulators in the home. In *2021 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1202–1209. doi: 10.1109 /RO-MAN50785.2021.9515511.

[30] Fanta Camara and Charles Fox. 2022. Extending quantitative proxemics and trust to hri. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 421–427. doi: 10.1109/RO-MAN53752 .2022.9900821.

[31] Tiffany Chen, Chih-Hung King, Andrea Thomaz, and Charles Kemp. 2014. An investigation of responses to robot-initiated touch in a nursing context. *International Journal of Social Robotics*, 6, (Jan. 2014). doi: 10.1007/s12369-013 -0215-x.

[32] Manolis Chiou, Faye McCabe, Markella Grigoriou, and Rustam Stolkin. 2021. Trust, shared understanding and locus of control in mixed-initiative robotic systems. In *2021 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 684–691. doi: 10.1109/RO-MAN50785 .2021.9515476.

[33] Jung Ju Choi, Yunkyung Kim, and Sonya S. Kwak. 2014. The autonomy levels and the human intervention levels of robots: the impact of robot types in human-robot interaction. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, 1069–1074. doi: 10.1109/ROMAN.201 4.6926394.

[34] John Christman. 2020. Autonomy in Moral and Political Philosophy. In *The Stanford Encyclopedia of Philosophy*. (Fall 2020 ed.). Edward N. Zalta, (Ed.) Metaphysics Research Lab, Stanford University.

[35] Jason R. Cody, Karina A. Roundtree, and Julie A. Adams. 2021. Human-collective collaborative target selection. *J. Hum.-Robot Interact.*, 10, 2, Article 18, (Mar. 2021), 29 pages. doi: 10.1145/3442679.

[36] sense 2.a cognition n. sense n. 2023. *Oxford English Dictionary*. Oxford University Press.

[37] Mihaela Constantinescu and Roger Crisp. 2022. Can robotic ai systems be virtuous and why does this matter? *International Journal of Social Robotics*, 14, 6, 1547–1557.

[38] Joseph Cooper and Michael A. Goodrich. 2008. Towards combining uav and sensor operator roles in uav-enabled visual search. In *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction* (HRI '08). Association for Computing Machinery, Amsterdam, The Netherlands, 351–358. isbn: 9781605580173. doi: 10.1145/1349822.1349868.

[39] Dagoberto Cruz-Sandoval, Arturo Morales-Tellez, Eduardo Benitez Sandoval, and Jesus Favela. 2020. A social robot as therapy facilitator in interventions to deal with dementia-related behavioral symptoms. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (HRI '20). Association for Computing Machinery, Cambridge, United Kingdom, 161–169. ISBN: 9781450367462. DOI: 10.1145/3319502.3374840.

[40] W.G.P. D. 2017. *The Promise of Human Autonomy*. Xlibris US. ISBN: 9781477127766. https://books.google.com/books?id=eycrDwAAQBAJ.

[41] Kate Darling. 2012. Extending legal protection to social robots: the effects of anthropomorphism, empathy, and violent behavior towards robotic objects. *SSRN Electronic Journal*, (Apr. 2012). DOI: 10.2139/ssrn.2044797.

[42] Devleena Das, Siddhartha Banerjee, and Sonia Chernova. 2021. Explainable ai for robot failures: generating explanations that improve user assistance in fault recovery. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction* (HRI '21). Association for Computing Machinery, Boulder, CO, USA, 351–360. ISBN: 9781450382892. DOI: 10.1145/3434073.34446 57.

[43] Amel Dechemi, Vikarn Bhakri, Ipsita Sahin, Arjun Modi, Julya Mestas, Pamodya Peiris, Dannya Enriquez Barrundia, Elena Kokkoni, and Konstantinos Karydis. 2021. Babynet: a lightweight network for infant reaching action recognition in unconstrained environments to support future pediatric rehabilitation applications. In *2021 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 461–467. DOI: 10.1109/RO-MAN50785 .2021.9515507.

[44] E.L. Deci and R. Flaste. 1995. *Why We Do what We Do: The Dynamics of Personal Autonomy. A Grosset/Putnam book*. Putnam's Sons. ISBN: 9780399140471. http s://books.google.com/books?id=gUnuAAAAMAAJ.

[45] Edward Deci and Richard Ryan. 2008. Self-determination theory: a macrotheory of human motivation, development, and health. *Canadian Psychology-psychologie Canadienne - CAN PSYCHOL-PSYCHOL CAN*, 49, (Aug. 2008). DOI: 10.1037/a0012801.

[46] Maartje de Graaf. 2016. An ethical evaluation of human–robot relationships. *International Journal of Social Robotics*, 8, (Aug. 2016). DOI: 10.1007/s12369-01 6-0368-5.

[47] Jesse de Pagter. 2023. From eu robotics and ai governance to hri research: implementing the ethics narrative. *International Journal of Social Robotics*, 1–15.

[48] Munjal Desai, Poornima Kaniarasu, Mikhail Medvedev, Aaron Steinfeld, and Holly Yanco. 2013. Impact of robot failures and feedback on real-time trust. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 251–258. DOI: 10.1109/HRI.2013.6483596.

[49] Daniela Doroftei, Tom De Vleeschauwer, Salvatore Lo Bue, Michaël Dewyn, Frik Vanderstraeten, and Geert De Cubber. 2021. Human-agent trust evaluation in a digital twin context. In *2021 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 203–207. DOI: 10.1109/RO-MAN50785.2021.9515445.

[50] Jane Dryden. 2023. Autonomy. https://iep.utm.edu/autonomy/. Accessed: 2023-07-24. (2023).

[51] Francesco Del Duchetto, Paul Baxter, and Marc Hanheide. 2019. Lindsey the tour guide robot - usage patterns in a museum long-term deployment. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1–8. DOI: 10.1109/RO-MAN46459.2019.8956329.

[52] Nicholas Epley, Adam Waytz, et al. 2010. Mind perception. *Handbook of social psychology*, 1, 5, 498–541.

[53] Reza Etemad-Sajadi, Antonin Soussan, and Théo Schöpfer. 2022. How ethical issues raised by human–robot interaction can impact the intention to use the robot? *International Journal of Social Robotics*, 14, (June 2022). DOI: 10.1007/s1 2369-021-00857-8.

[54] Haruhiko Eto and H. Harry Asada. 2019. Seamless manual-to-autopilot transition: an intuitive programming approach to robotic welding. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1–7. DOI: 10.1109/RO-MAN46459.2019.8956403.

[55] Elisa Foderaro, Amedeo Cesta, Alessandro Umbrico, and Andrea Orlandini. 2021. Simplifying the a.i. planning modeling for human-robot collaboration. In *2021 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1011–1016. DOI: 10.1109/RO-MAN50785.2021.951 5431.

[56] Deen K. Chatterjee, (Ed.) 2011. *Political autonomy. Encyclopedia of Global Justice*. Springer Netherlands, Dordrecht, 853–854. ISBN: 978-1-4020-9160-5. DOI: 10.1007/978-1-4020-9160-5_350.

[57] Matthew C. Fontaine and Stefanos Nikolaidis. 2022. Evaluating human–robot interaction algorithms in shared autonomy via quality diversity scenario generation. *J. Hum.-Robot Interact.*, 11, 3, Article 25, (Sept. 2022), 30 pages. DOI: 10.1145/3476412.

[58] Eduard Fosch Villaronga, Pranav Khanna, Hadassah Drukarch, and Bart Custers. 2022. The role of humans in surgery automation: exploring the influence of automation on human–robot interaction and responsibility in surgery innovation. *International Journal of Social Robotics*, 15, (Apr. 2022). DOI: 10.1007/s12369-022-00875-0.

[59] Harry Frankfurt. 1999. On caring. In *Necessity, Volition, and Love*. Cambridge University Press, Cambridge. Chap. 10, 155–180.

[60] Denise Y. Geiskkovitch, Derek Cormier, Stela H. Seo, and James E. Young. 2016. Please continue, we need more data: an exploration of obedience to robots. *J. Hum.-Robot Interact.*, 5, 1, (Mar. 2016), 82–99.

[61] Ken Gemes and Christopher Janaway. 2006. Nietzsche on free will, autonomy and the sovereign individual. *Aristotelian Society Supplementary Volume*, 80, 1, 321–357. DOI: 10.1111/j.0066-7373.2006.00135.x.

[62] Soheil Gholami, Virginia Ruiz Garate, Elena De Momi, and Arash Ajoudani. 2020. A shared-autonomy approach to goal detection and navigation control of mobile collaborative robots. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1026–1032. DOI: 10.1109/RO-MAN47096.2020.9223583.

[63] R Gillon. 1985. Autonomy and the principle of respect for autonomy. *BMJ*, 290, 6484, 1806–1808. eprint: https://www.bmj.com/content/290/6484/1806.fu ll.pdf. DOI: 10.1136/bmj.290.6484.1806.

[64] Dylan F. Glas, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2009. Field trial for simultaneous teleoperation of mobile social robots. In *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction* (HRI '09). Association for Computing Machinery, La Jolla, California, USA, 149–156. ISBN: 9781605584041. DOI: 10.1145/1514095.1514123.

[65] Eduardo González and Fernando A. Auat Cheein. 2018. Preliminary results on reducing the workload of assistive vehicle users: a collaborative driving approach. *International Journal of Social Robotics*, 10, 555–568.

[66] John-Stewart Gordon. 2023. Bioethics. https://iep.utm.edu/bioethics/. Accessed: 2023-07-24. (2023).

[67] Arzu Guneysu and Bert Arnrich. 2017. Socially assistive child-robot interaction in physical exercise coaching. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 670–675. DOI: 10.1109/ROMAN.2017.8172375.

[68] Yue Guo, Boshi Wang, Dana Hughes, Michael Lewis, and Katia Sycara. 2020. Designing context-sensitive norm inverse reinforcement learning framework for norm-compliant autonomous agents. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 618–625. DOI: 10.1109/RO-MAN47096.2020.9223344.

[69] Paul Guyer. 2003. Kant on the theory and practice of autonomy. *Social Philosophy and Policy*, 20, 2, 70–98. DOI: 10.1017/S026505250320203X.

[70] Belaidi Hadjira, Abdelfetah Hentout, and Hamid Bentarzi. 2019. Human–robot shared control for path generation and execution. *International Journal of Social Robotics*, 11, (Aug. 2019). DOI: 10.1007/s12369-019-00520-3.

[71] John Hadley. 2017. Non-autonomous sentient beings and original acquisition. *Analysis*, 77, 2, pp. 293–299. Retrieved June 13, 2023 from https://www.jstor.o rg/stable/48545185.

[72] Janik Hager, Ruben Bauer, Marc Toussaint, and Jim Mainprice. 2021. Graspme - grasp manifold estimator. In *2021 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 626–632. DOI: 10.1109/RO-MAN50785.2021.9515479.

[73] Marc Hanheide, Denise Hebesberger, and Tomáš Krajník. 2017. The when, where, and how: an adaptive robotic info-terminal for care home residents. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (HRI '17). Association for Computing Machinery, Vienna, Austria, 341–349. ISBN: 9781450343367. DOI: 10.1145/2909824.3020228.

[74] 2023. Hanson robotics sophia. https://www.hansonrobotics.com/sophia/. Accessed: 2023-07-24. (2023).

[75] Benjamin Hardin and Michael A. Goodrich. 2009. On using mixed-initiative control: a perspective for managing large-scale robotic teams. In *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction* (HRI '09). Association for Computing Machinery, La Jolla, California, USA, 165–172. ISBN: 9781605584041. DOI: 10.1145/1514095.1514126.

[76] Jamie Harris and Jacy Reese Anthis. 2021. The Moral Consideration of Artificial Entities: A Literature Review. *Science and Engineering Ethics*, 27, 4, (Aug. 2021), 53. DOI: 10.1007/s11948-021-00331-8.

[77] Vahagn Harutyunyan, Vimitha Manohar, Issak Gezehei, and Jacob W. Crandall. 2013. Cognitive telepresence in human-robot interactions. *J. Hum.-Robot Interact.*, 1, 2, (Jan. 2013), 158–182. DOI: 10.5898/JHRI.1.2.Harutyunyan.

[78] Denise Hebesberger, Tobias Koertner, Christoph Gisinger, and Jürgen Pripfl. 2017. A long-term autonomous robot at a care hospital: a mixed methods study on social acceptance and experiences of staff and older adults. *International Journal of Social Robotics*, 9, 3, 417–429.

[79] Ala'aldin Hijaz, Jessica Korneder, and Wing-Yue Geoffrey Louie. 2021. In-the-wild learning from demonstration for therapies for autism spectrum disorder. In *2021 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1224–1229. DOI: 10.1109/RO-MAN50785.2021.951 5439.

[80] Ala'aldin Hijaz, Wing-Yue Geoffrey Louie, and Iyad Mansour. 2019. Towards a driver monitoring system for estimating driver situational awareness. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1–6. DOI: 10.1109/RO-MAN46459.2019.8956378.

[81] Junya Hirose, Masakazu Hirokawa, and Kenji Suzuki. 2014. Robotic gaming companion to facilitate social interaction among children. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, 63–68. DOI: 10.1109/ROMAN.2014.6926231.

[82] Toshio Hori and Yoshifumi Nishida. 2008. Improvement of position estimation of the ultrasonic 3d tag system. In *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*, 436–441. DOI: 10.1109/ROMAN.2008.4600705.

[83] Hanns-Peter Horn, Matthias Nadig, Johannes Hackbarth, Christian Willms, Caspar Jacob, Serge Autexier, Tim Schwartz, and Ivana Kruijff-Korbayová. 2022. Iterative user-centric development of mobile robotic systems with intuitive multimodal human-robot interaction in a clinic environment. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 963–968. DOI: 10.1109/RO-MAN53752.2022.9900703.

[84] Aike C. Horstmann and Nicole C. Krämer. 2022. The fundamental attribution error in human-robot interaction: an experimental investigation on attributing responsibility to a social robot for its pre-programmed behavior. *International Journal of Social Robotics*, 14, 5, 1137–1153.

[85] Hui-Min Huang. 2007. Autonomy levels for unmanned systems (alfus) framework: safety and application issues. In (PerMIS '07). Association for Computing Machinery, Washington, D.C., 48–53. ISBN: 9781595938541. DOI: 10.1145/1660877.1660883.

[86] Sophie Husemann, Jan Pöppel, and Stefan Kopp. 2022. Differences and biases in mentalizing about humans and robots. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 490–497. DOI: 10.1109/RO-MAN53752.2022.9900849.

[87] Rainer Jäkel, Sven R. Schmidt-Rohr, Steffen W. Rühl, Alexander Kasper, Zhixing Xue, and Rüdiger Dillmann. 2012. Learning of planning models for dexterous manipulation based on human demonstrations. *International Journal of Social Robotics*, 4, 437–448.

[88] Petar Jerčić, Wei Wen, Johan Hagelbäck, and Veronica Sundstedt. 2018. The effect of emotions and social behavior on performance in a collaborative serious game between humans and autonomous robots. *International Journal of Social Robotics*, 10, (Jan. 2018), 1–15. DOI: 10.1007/s12369-017-0437-4.

[89] Michelle J. Johnson, Mayumi Mohan, and Rochelle Mendonca. 2022. Therapist-patient interactions in task-oriented stroke therapy can guide robot-patient interactions. *International Journal of Social Robotics*, 14, 6, 1527–1546.

[90] Robert Johnson and Adam Cureton. 2022. Kant's Moral Philosophy. In *The Stanford Encyclopedia of Philosophy*. (Fall 2022 ed.). Edward N. Zalta and Uri Nodelman, (Eds.) Metaphysics Research Lab, Stanford University.

[91] Immanuel Kant. 1998. *Groundwork of the Metaphysics of Morals*. Mary Gregor, (Ed.) Cambridge University Press, Cambridge.

[92] Richard Kelley, Alireza Tavakkoli, Christopher King, Monica Nicolescu, Mircea Nicolescu, and George Bebis. 2008. Understanding human intentions via hidden markov models in autonomous mobile robots. In *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction* (HRI '08). Association for Computing Machinery, Amsterdam, The Netherlands, 367–374. ISBN: 9781605580173. DOI: 10.1145/1349822.1349870.

[93] Taenyun Kim, Maria D. Molina, Minjin (MJ) Rheu, Emily S. Zhan, and Wei Peng. 2023. One ai does not fit all: a cluster analysis of the laypeople's perception of ai roles. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (CHI '23) Article 29. Association for Computing Machinery, Hamburg, Germany, 20 pages. ISBN: 9781450394215. DOI: 10.1145/3544548.3581340.

[94] Takuya Kitade, Satoru Satake, Takayuki Kanda, and Michita Imai. 2013. Understanding suitable locations for waiting. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 57–64. DOI: 10.1109/HRI.2013.6483502.

[95] Andreas Kolling, Steven Nunnally, and Michael Lewis. 2012. Towards human control of robot swarms. In *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction* (HRI '12). Association for Computing Machinery, Boston, Massachusetts, USA, 89–96. ISBN: 9781450310635. DOI: 10.1145/2157689.2157704.

[96] Elly Konijn, Brechtje Jansen, Victoria Bustos, Veerle Hobbelink, and Daniel Vanegas. 2022. Social robots for (second) language learning in (migrant) primary school children. *International Journal of Social Robotics*, 14, (Apr. 2022), 1–17. DOI: 10.1007/s12369-021-00824-3.

[97] J. E. (Hans). Korteling, G. C. van de Boer-Visschedijk, R. A. M. Blankendaal, R. C. Boonekamp, and A. R. Eikelboom. 2021. Human- versus artificial intelligence. *Frontiers in Artificial Intelligence*, 4. DOI: 10.3389/frai.2021.622364.

[98] Andrea Krausman et al. 2022. Trust measurement in human-autonomy teams: development of a conceptual toolkit. *J. Hum.-Robot Interact.*, 11, 3, Article 33, (Sept. 2022), 58 pages. DOI: 10.1145/3530874.

[99] Jun Ki Lee, Robert Lopez Toscano, Walter Dan Stiehl, and Cynthia Breazeal. 2008. The design of a semi-autonomous robot avatar for family communication and education. In *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*, 166–173. DOI: 10.1109/ROMAN.2008.4600661.

[100] Florent Levillain and Elisabetta Zibetti. 2017. Behavioral objects: the rise of the evocative machines. *J. Hum.-Robot Interact.*, 6, 1, (May 2017), 4–24. DOI: 10.5898/JHRI.6.1.Levillain.

[101] Jing Li, Hua Peng, Huosheng Hu, Zhiming Luo, and Chao Tang. 2020. Multimodal information fusion for automatic aesthetics evaluation of robotic dance poses. *International Journal of Social Robotics*, 12, (Jan. 2020). DOI: 10.1007/s12369-019-00535-w.

[102] Yuhua Liang and Seungcheol Lee. 2017. Fear of autonomous robots and artificial intelligence: evidence from national representative data with probability sampling. *International Journal of Social Robotics*, 9, (June 2017). DOI: 10.1007/s12369-017-0401-3.

[103] Felix Lindner. 2016. A model of a robot's will based on higher-order desires. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 749–754. DOI: 10.1109/ROMAN.2016.7745203.

[104] Alexandru Litoiu, Daniel Ullman, Jason Kim, and Brian Scassellati. 2015. Evidence that robots trigger a cheating detector in humans. In *Proceedings of the tenth annual acm/ieee international conference on human-robot interaction*, 165–172.

[105] Maria Mannone, Valeria Seidita, and Antonio Chella. 2022. Quantum robosound: auditory feedback of a quantum-driven robotic swarm. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 287–292. DOI: 10.1109/RO-MAN53752.2022.9900578.

[106] Frank Martela and Tapani J. J. Riekki. 2018. Autonomy, competence, relatedness, and beneficence: a multicultural comparison of the four pathways to meaningful work. *Frontiers in Psychology*, 9. DOI: 10.3389/fpsyg.2018.01157.

[107] Marcus Mast, Zdeněk Materna, Michal Španěl, Florian Weisshardt, Georg Arbeiter, Michael Burmester, Pavel Smrž, and Birgit Graf. 2014. Semi-autonomous domestic service robots: evaluation of a user interface for remote manipulation and navigation with focus on effects of stereoscopic display. *International Journal of Social Robotics*, 7, (Apr. 2014), 183–202. DOI: 10.1007/s12369-014-0266-7.

[108] Marcus Mast et al. 2012. User-centered design of a dynamic-autonomy remote interaction concept for manipulation-capable robots to assist elderly people in the home. *J. Hum.-Robot Interact.*, 1, 1, (July 2012), 96–118. DOI: 10.5898/JHRI.1.1.Mast.

[109] Christoforos Mavrogiannis, Alena M. Hutchinson, John Macdonald, Patrícia Alves-Oliveira, and Ross A. Knepper. 2019. Effects of distinct robot navigation strategies on human behavior in a crowded environment. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 421–430. DOI: 10.1109/HRI.2019.8673115.

[110] Samuel J. McDonald, Mark B. Colton, C. Kristopher Alder, and Michael A. Goodrich. 2017. Haptic shape-based management of robot teams in cordon and patrol. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (HRI '17). Association for Computing Machinery, Vienna, Austria, 380–388. ISBN: 9781450343367. DOI: 10.1145/2909824.3020243.

[111] Lingheng Meng, Daiwei Lin, Adam Francey, Rob Gorbet, Philip Beesley, and Dana Kulić. 2020. Learning to engage with interactive systems: a field study on deep reinforcement learning in a public museum. *J. Hum.-Robot Interact.*, 10, 1, Article 5, (Oct. 2020), 29 pages. DOI: 10.1145/3408876.

[112] Rajesh Elara Mohan, Wijerupage Wijesoma, Carlos Calderon, and Changjiu Zhou. 2009. Experimenting false alarm demand for human robot interactions in humanoid soccer robots. *I. J. Social Robotics*, 1, (Apr. 2009), 171–180. DOI: 10.1007/s12369-009-0017-3.

[113] AJung Moon, Peter Danielson, and HF Machiel Van der Loos. 2012. Survey-based discussions on morally contentious applications of interactive robotics. *International Journal of Social Robotics*, 4, 77–96.

[114] Luis Yoichi Morales Saiki, Satoru Satake, Takayuki Kanda, and Norihiro Hagita. 2011. Modeling environments from a route perspective. In *Proceedings of the 6th International Conference on Human-Robot Interaction* (HRI '11). Association for Computing Machinery, Lausanne, Switzerland, 441–448. ISBN: 9781450305617. DOI: 10.1145/1957656.1957815.

[115] Luca Morando, Carmine Tommaso Recchiuto, and Antonio Sgorbissa. 2020. Social drone sharing to increase the uav patrolling autonomy in emergency scenarios. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 539–546. DOI: 10.1109/RO-MAN47096.2020.9223567.

[116] Christopher E. Mower, João Moura, Aled Davies, and Sethu Vijayakumar. 2019. Modulating human input for shared autonomy in dynamic environments. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1–8. DOI: 10.1109/RO-MAN46459.2019.8956304.

[117] Pierre-André Mudry, Sarah Degallier, and Aude Billard. 2008. On the influence of symbols and myths in the responsibility ascription problem in roboethics-a roboticist's perspective. In *RO-MAN 2008-The 17th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 563–568.

[118] Amir M. Naghsh, Jeremi Gancet, Andry Tanoto, and Chris Roast. 2008. Analysis and design of human-robot swarm interaction in firefighting. In *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*, 255–260. DOI: 10.1109/ROMAN.2008.4600675.
[119] Junya Nakanishi, Tomohisa Hazama, Jun Baba, Sichao Song, Yuichiro Yoshikawa, and Hiroshi Ishiguro. 2021. Exploring possibilities of social robot's interactive services in the case of a hotel room. In *2021 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 925–930. DOI: 10.1109/RO-MAN50785.2021.9515380.
[120] Amal Nanavati, Nick Walker, Lee Taber, Christoforos Mavrogiannis, Leila Takayama, Maya Cakmak, and Siddhartha Srinivasa. 2022. Not all who wander are lost: a localization-free system for in-the-wild mobile robot deployments. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 422–431. DOI: 10.1109/HRI53351.2022.9889620.
[121] Jauwairia Nasir, Pierre Oppliger, Barbara Bruno, and Pierre Dillenbourg. 2022. Questioning wizard of oz: effects of revealing the wizard behind the robot. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1385–1392. DOI: 10.1109/RO-MAN53752.2022.9900718.
[122] Clifford Nass, Jonathan Steuer, and Ellen R. Tauber. 1994. Computers are social actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '94). Association for Computing Machinery, Boston, Massachusetts, USA, 72–78. ISBN: 0897916506. DOI: 10.1145/191666.191703.
[123] Simone Nertinger, Robin Kirschner, Djallil Naceri, and Sami Haddadin. 2022. Acceptance of remote assistive robots with and without human-in-the-loop for healthcare applications. *International Journal of Social Robotics*, (Oct. 2022). DOI: 10.1007/s12369-022-00931-9.
[124] Stefanos Nikolaidis, Yu Xiang Zhu, David Hsu, and Siddhartha Srinivasa. 2017. Human-robot mutual adaptation in shared autonomy. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (HRI '17). Association for Computing Machinery, Vienna, Austria, 294–302. ISBN: 9781450343367. DOI: 10.1145/2909824.3020252.
[125] Timothy O'Connor and Christopher Franklin. 2022. Free Will. In *The Stanford Encyclopedia of Philosophy*. (Winter 2022 ed.). Edward N. Zalta and Uri Nodelman, (Eds.) Metaphysics Research Lab, Stanford University.
[126] Yoojin Oh, Tim Schäfer, Benedikt Rüther, Marc Toussaint, and Jim Mainprice. 2021. A system for traded control teleoperation of manipulation tasks using intent prediction from hand gestures. In *2021 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 503–508. DOI: 10.1109/RO-MAN50785.2021.9515440.
[127] Yoojin Oh, Shao-Wen Wu, Marc Toussaint, and Jim Mainprice. 2020. Natural gradient shared control. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1223–1229. DOI: 10.1109/RO-MAN47096.2020.9223465.
[128] Suman Ojha, Jonathan Vitale, and Mary-Anne Williams. 2021. Computational emotion models: a thematic review. *International Journal of Social Robotics*, 13, (Sept. 2021). DOI: 10.1007/s12369-020-00713-1.
[129] Suman Ojha, Mary-Anne Williams, and Benjamin Johnston. 2018. The essence of ethical reasoning in robot-emotion processing. *International Journal of Social Robotics*, 10, 211–223.
[130] Kazuo Okamura and Seiji Yamada. 2020. Calibrating trust in human-drone cooperative navigation. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1274–1279. DOI: 10.1109/RO-MAN47096.2020.9223509.
[131] Samuel A. Olatunji, Andre Potenza, Andrey Kiselev, Tal Oron-Gilad, Amy Loutfi, and Yael Edan. 2022. Levels of automation for a mobile robot teleoperated by a caregiver. *J. Hum.-Robot Interact.*, 11, 2, Article 20, (Feb. 2022), 21 pages. DOI: 10.1145/3507471.
[132] Masahiko Osawa, Kohei Okuoka, Yusuke Takimoto, and Michita Imai. 2020. Is automation appropriate? semi-autonomous telepresence architecture focusing on voluntary and involuntary movements. *International Journal of Social Robotics*, 12, 5, 1119–1134.
[133] Anastasia K. Ostrowski, Raechel Walker, Madhurima Das, Maria Yang, Cynthia Breazeal, Hae Won Park, and Aditi Verma. 2022. Ethics, equity, & justice in human-robot interaction: a review and future directions. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 969–976. DOI: 10.1109/RO-MAN53752.2022.9900805.
[134] Anita Paas, Emily B. J. Coffey, Giovanni Beltrame, and David St-Onge. 2022. Towards evaluating the impact of swarm robotic control strategy on operators' cognitive load. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 217–223. DOI: 10.1109/RO-MAN53752.2022.9900763.
[135] Amit Kumar Pandey and Rachid Alami. 2014. Towards human-level semantics understanding of human-centered object manipulation tasks for hri: reasoning about effect, ability, effort and perspective taking. *International Journal of Social Robotics*, 6, 593–620.
[136] Dario Pasquali, Jonas Gonzalez Billandon, Alexander Aroyo, Giulio Sandini, Alessandra Sciutti, and Francesco Rea. 2021. Detecting lies is a child (robot)'s play: gaze-based lie detection in hri. *International Journal of Social Robotics*, 15, (Nov. 2021). DOI: 10.1007/s12369-021-00822-5.
[137] Janet V.T. Pauketat and Jacy Reese Anthis. 2022. Predicting the moral consideration of artificial intelligences. *Computers in Human Behavior*, 136, (Nov. 2022), 107372. DOI: 10.1016/j.chb.2022.107372.
[138] Xavier Perrin, Ricardo Chavarriaga, Céline Ray, Roland Siegwart, and José del R. Millán. 2008. A comparative psychophysical and eeg study of different feedback modalities for hri. In *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction* (HRI '08). Association for Computing Machinery, Amsterdam, The Netherlands, 41–48. ISBN: 9781605580173. DOI: 10.1145/1349822.1349829.
[139] Frano Petric and Zdenko Kovačić. 2019. Hierarchical pomdp framework for a robot-assisted asd diagnostic protocol. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 286–293. DOI: 10.1109/HRI.2019.8673295.
[140] José Carlos Pulido, José Carlos González Dorado, Cristina Suárez, Antonio Bandera, Pablo Bustos, and Fernando Fernández. 2017. Evaluating the child–robot interaction of the naotherapist platform in pediatric rehabilitation. *International Journal of Social Robotics*, 9, (June 2017), 343–358. DOI: 10.1007/s12369-017-0402-2.
[141] Nicole L. Robinson, Jennifer Connolly, Gavin Suddery, Madeliene Turner, and David J. Kavanagh. 2021. A humanoid social robot to provide personalized feedback for health promotion in diet, physical activity, alcohol and cigarette use: a health clinic trial. In *2021 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 720–726. DOI: 10.1109/RO-MAN50785.2021.9515558.
[142] Marta Romeo, Angelo Cangelosi, and Ray Jones. 2018. Developing a deep learning agent for hri: dataset collection and training. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 1150–1155. DOI: 10.1109/ROMAN.2018.8525512.
[143] Astrid Rosenthal-von der Pütten, Carolin Straßmann, and Martina Mara. 2017. A long time ago in a galaxy far, far away…the effects of narration and appearance on the perception of robots. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 1169–1174. DOI: 10.1109/ROMAN.2017.8172452.
[144] Silvia Rossi, Alessandra Rossi, and Kerstin Dautenhahn. 2020. The secret life of robots: perspectives and challenges for robot's behaviours during non-interactive tasks. *International Journal of Social Robotics*, 12, (Dec. 2020). DOI: 10.1007/s12369-020-00650-z.
[145] Aaron M. Roth, Samantha Reig, Umang Bhatt, Jonathan Shulgach, Tamara Amin, Afsaneh Doryab, Fei Fang, and Manuela Veloso. 2019. A robot's expressive language affects human strategy and perceptions in a competitive game. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1–8. DOI: 10.1109/RO-MAN46459.2019.8956412.
[146] Sayanti Roy, Trey Smith, Brian Coltin, and Tom Williams. 2023. I need your help... or do i? maintaining situation awareness through performative autonomy. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction* (HRI '23). Association for Computing Machinery, Stockholm, Sweden, 122–131. ISBN: 9781450399647. DOI: 10.1145/3568162.3576954.
[147] Matthew Rueben, Jeffrey Klow, Madelyn Duer, Eric Zimmerman, Jennifer Piacentini, Madison Browning, Frank J. Bernieri, Cindy M. Grimm, and William D. Smart. 2021. Mental models of a mobile shoe rack: exploratory findings from a long-term in-the-wild study. *J. Hum.-Robot Interact.*, 10, 2, Article 16, (Feb. 2021), 36 pages. DOI: 10.1145/3442620.
[148] Adam Rule and Jodi Forlizzi. 2012. Designing interfaces for multi-user, multi-robot systems. In *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction* (HRI '12). Association for Computing Machinery, Boston, Massachusetts, USA, 97–104. ISBN: 9781450310635. DOI: 10.1145/2157689.2157705.
[149] Richard Ryan and Edward Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American psychologist*, 55, (Feb. 2000), 68–78. DOI: 10.1037/0003-066X.55.1.68.
[150] Tamie Salter, François Michaud, and Hélène Larouche. 2010. How wild is wild? a taxonomy to characterize the 'wildness' of child-robot interaction. *I. J. Social Robotics*, 2, (Dec. 2010), 405–415. DOI: 10.1007/s12369-010-0069-4.
[151] Marcus M. Scheunemann, Raymond H. Cuijpers, and Christoph Salge. 2020. Warmth and competence to predict human preference of robot behavior in physical human-robot interaction. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1340–1347. DOI: 10.1109/RO-MAN47096.2020.9223478.
[152] Markus Schlosser. 2019. Agency. In *The Stanford Encyclopedia of Philosophy*. (Winter 2019 ed.). Edward N. Zalta, (Ed.) Metaphysics Research Lab, Stanford University.
[153] Melanie Schmidt-Wolf and David Feil-Seifer. 2022. Comparison of vehicle-to-bicyclist and vehicle-to-pedestrian communication feedback module: a study on increasing legibility, public acceptance and trust. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1058–1064. DOI: 10.1109/RO-MAN53752.2022.9900748.

[154] Sebastian Schneider and Franz Kummert. 2021. Comparing robot and human guided personalization: adaptive exercise robots are perceived as more competent and trustworthy. *International Journal of Social Robotics*, 13, (Apr. 2021). doi: 10.1007/s12369-020-00629-w.

[155] Thomas B. Sheridan, William L. Verplank, and T. L. Brooks. 1978. Human/computer control of undersea teleoperators. In *NASA. Ames Res. Center The 14th Ann. Conf. on Manual Control*.

[156] Masahiro Shiomi, Daisuke Sakamoto, Takayuki Kanda, Carlos Toshinori Ishi, Hiroshi Ishiguro, and Norihiro Hagita. 2008. A semi-autonomous communication robot: a field trial at a train station. In *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction* (HRI '08). Association for Computing Machinery, Amsterdam, The Netherlands, 303–310. isbn: 9781605580173. doi: 10.1145/1349822.1349862.

[157] Rosanne M. Siino, Justin Chung, and Pamela J. Hinds. 2008. Colleague vs. tool: effects of disclosure in human-robot collaboration. In *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*, 558–562. doi: 10.1109/ROMAN.2008.4600725.

[158] Grimaldo Silva, Khansa Rekik, Ali Kanso, and Leizer Schnitman. 2022. Multi-perspective human robot interaction through an augmented video interface supported by deep learning. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1168–1173. doi: 10.1109/RO-MAN53752.2022.9900671.

[159] Tim Smithers. 1997. Autonomy in robots and other agents. *Brain and Cognition*, 34, 1, 88–106. doi: https://doi.org/10.1006/brcg.1997.0908.

[160] Julia Stapels and Friederike Eyssel. 2022. Robocalypse? yes, please! the role of robot autonomy in the development of ambivalent attitudes towards robots. *International Journal of Social Robotics*, 14, (Apr. 2022). doi: 10.1007/s12369-021-00817-2.

[161] Halit Bener Suay, Russell Toris, and Sonia Chernova. 2012. A practical comparison of three robot learning from demonstration algorithm. *International Journal of Social Robotics*, 4, 319–330.

[162] Tao Sun, Donghao Shao, Zhendong Dai, and Poramate Manoonpong. 2018. Adaptive neural control for self-organized locomotion and obstacle negotiation of quadruped robots. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 1081–1086. doi: 10.1109/ROMAN.2018.8525645.

[163] Kazuaki Tanaka, Naomi Yamashita, Hideyuki Nakanishi, and Hiroshi Ishiguro. 2016. Teleoperated or autonomous?: how to produce a robot operator's pseudo presence in hri. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 133–140. doi: 10.1109/HRI.2016.7451744.

[164] 2021. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. https://www.sae.org/standards/content/j3016_202104/. Accessed: 2023-07-24. (Apr. 2021).

[165] Michael Thompson. 2017. Autonomy and common good: interpreting rousseau's general will. *International Journal of Philosophical Studies*, 25, (Feb. 2017), 1–20. doi: 10.1080/09672559.2017.1286364.

[166] Suzanne Tolmeijer, Astrid Weiss, Marc Hanheide, Felix Lindner, Thomas M. Powers, Clare Dixon, and Myrthe L. Tielman. 2020. Taxonomy of trust-relevant failures and mitigation strategies. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (HRI '20). Association for Computing Machinery, Cambridge, United Kingdom, 3–12. isbn: 9781450367462. doi: 10.1145/3319502.3374793.

[167] Ilaria Torre, Alexis Linard, Anders Steen, Jana Tumová, and Iolanda Leite. 2021. Should robots chicken? how anthropomorphism and perceived autonomy influence trajectories in a game-theoretic problem. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction* (HRI '21). Association for Computing Machinery, Boulder, CO, USA, 370–379. isbn: 9781450382892. doi: 10.1145/3434073.3444687.

[168] Daniel Tozadore, Adam Pinto, Roseli Romero, and Gabriele Trovato. 2017. Wizard of oz vs autonomous: children's perception changes according to robot's operation condition. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 664–669. doi: 10.1109/ROMAN.2017.8172374.

[169] Alexander Tyshka and Wing-Yue Geoffrey Louie. 2022. Transparent learning from demonstration for robot-mediated therapy. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 891–897. doi: 10.1109/RO-MAN53752.2022.9900854.

[170] Daniel Ullman, Salomi Aladia, and Bertram F. Malle. 2021. Challenges and opportunities for replication science in hri: a case study in human-robot trust. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction* (HRI '21). Association for Computing Machinery, Boulder, CO, USA, 110–118. isbn: 9781450382892. doi: 10.1145/3434073.3444652.

[171] Steven Umbrello and Roman Yampolskiy. 2022. Designing ai for explainability and verifiability: a value sensitive design approach to avoid artificial stupidity in autonomous vehicles. *International Journal of Social Robotics*, 14, (Mar. 2022), 1–10. doi: 10.1007/s12369-021-00790-w.

[172] Guillaume Vailland, Yoren Gaffary, Louise Devigne, Valérie Gouranton, Bruno Arnaldi, and Marie Babel. 2020. Vestibular feedback on a virtual reality wheelchair driving simulator: a pilot study. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (HRI '20). Association for Computing Machinery, Cambridge, United Kingdom, 171–179. isbn: 9781450367462. doi: 10.1145/3319502.3374825.

[173] Alberto Valero, Massimo Mecella, Fernando Matia, and Daniele Nardi. 2008. Adaptative human-robot interaction for mobile robots. In *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*, 243–248. doi: 10.1109/ROMAN.2008.4600673.

[174] Dieter Vanderelst and Jurgen Willems. 2020. Can we agree on what robots should be allowed to do? an exercise in rule selection for ethical care robots. *International Journal of Social Robotics*, 12, 1093–1102.

[175] Josca van Houwelingen-Snippe, Jered Vroon, Gwenn Englebienne, and Pim Haselager. 2017. Blame my telepresence robot joint effect of proxemics and attribution on interpersonal attraction. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 162–168. doi: 10.1109/ROMAN.2017.8172296.

[176] Andrea Vanzo, Francesco Riccio, Mahmoud Sharf, Valeria Mirabella, Tiziana Catarci, and Daniele Nardi. 2020. Who is willing to help robots? a user study on collaboration attitude. *International Journal of Social Robotics*, 12, 589–598.

[177] Somogy Varga and Charles Guignon. 2023. Authenticity. In *The Stanford Encyclopedia of Philosophy*. (Summer 2023 ed.). Edward N. Zalta and Uri Nodelman, (Eds.) Metaphysics Research Lab, Stanford University.

[178] Vighnesh Vatsal and Guy Hoffman. 2017. Wearing your arm on your sleeve: studying usage contexts for a wearable robotic forearm. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 974–980. doi: 10.1109/ROMAN.2017.8172421.

[179] Manuela M. Veloso. 2018. The increasingly fascinating opportunity for human-robot-ai interaction: the cobot mobile service robots. *J. Hum.-Robot Interact.*, 7, 1, Article 5, (May 2018), 2 pages. doi: 10.1145/3209541.

[180] Huadong Wang et al. 2011. Scalable target detection for large robot teams. In *Proceedings of the 6th International Conference on Human-Robot Interaction* (HRI '11). Association for Computing Machinery, Lausanne, Switzerland, 363–370. isbn: 9781450305617. doi: 10.1145/1957656.1957792.

[181] Yueh-Hsuan Weng, Chien-Hsun Chen, and Chuen-Tsai Sun. 2009. Toward the human–robot co-existence society: on safety intelligence for next generation robots. *International Journal of Social Robotics*, 1, 267–282.

[182] Eva Wiese, Giorgio Metta, and Agnieszka Wykowska. 2017. Robots as intentional agents: using neuroscientific methods to make robots appear more social. *Frontiers in Psychology*, 8. doi: 10.3389/fpsyg.2017.01663.

[183] Jason R. Wilson, Linda Tickle-Degnen, and Matthias Scheutz. 2020. Challenges in designing a fully autonomous socially assistive robot for people with parkinson's disease. *J. Hum.-Robot Interact.*, 9, 3, Article 20, (May 2020), 31 pages. doi: 10.1145/3379179.

[184] Susan Wolf. 1987. Sanity and the metaphysics of responsibility. In *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*. Ferdinand David Schoeman, (Ed.) Cambridge University Press, 46–62.

[185] Luke Wood, Abolfazl Zaraki, Ben Robins, and Kerstin Dautenhahn. 2021. Developing kaspar: a humanoid robot for children with autism. *International Journal of Social Robotics*, 13, (June 2021). doi: 10.1007/s12369-019-00563-6.

[186] Lauren L. Wright, Aditi Kothiyal, Kai O. Arras, and Barbara Bruno. 2022. How a social robot's vocalization affects children's speech, learning, and interaction. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 279–286. doi: 10.1109/RO-MAN53752.2022.9900811.

[187] Guang-Zhong Yang et al. 2017. Medical robotics–regulatory, ethical, and legal considerations for increasing levels of autonomy. *Science Robotics*, 2, 4, eaam8638. eprint: https://www.science.org/doi/pdf/10.1126/scirobotics.aam8638. doi: 10.1126/scirobotics.aam8638.

[188] Setareh Zafari and Sabine T. Koeszegi. 2021. Attitudes toward attributed agency: role of perceived control. *International Journal of Social Robotics*, 13, 8, 2071–2080.

[189] Abolfazl Zaraki, Luke Wood, Ben Robins, and Kerstin Dautenhahn. 2018. Development of a semi-autonomous robotic system to assist children with autism in developing visual perspective taking skills. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 969–976. doi: 10.1109/ROMAN.2018.8525681.

[190] Zhizheng Zhang, Zhibo Chen, and Weiping Li. 2018. Automating robotic furniture with a collaborative vision-based sensing scheme. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 719–725. doi: 10.1109/ROMAN.2018.8525783.

[191] Jakub Złotowski, Kumar Yogeeswaran, and Christoph Bartneck. 2017. Can we control it? autonomous robots threaten human identity, uniqueness, safety, and resources. *International Journal of Human-Computer Studies*, 100, 48–54. doi: https://doi.org/10.1016/j.ijhcs.2016.12.008.