The Psychology of Robots and Artificial Intelligence

Kurt Gray,1 Kai Chi Yam,2 Alexander Eng Zhen'An,2 Danica Wilbanks,1 Adam Waytz3

1. University of North Carolina, Department of Psychology and Neuroscience
2. National University of Singapore, Business School, Department of Management
3. Northwestern University, Kellogg School of Management, Department of Management and Operations

Accepted, Handbook of Social Psychology


Note: This is an excerpt of the full chapter, containing the introduction and the conclusion. It focuses on the idea of “replacement”. Note that the final chapter will likely change from current version, citing more (emerging) research on LLMs

Corresponding author:
Kurt Gray
CB 3270
University of North Carolina at Chapel Hill
Chapel Hill, NC 27510
kurtjgray@gmail.com
### Table of Contents

**Introduction to The Psychology of Robots and Artificial Intelligence** ........ 2  
**Machines as Agents of Replacement** ........................................... 4  
AI and Robots as Agents of Replacement 5  
**The Mind of Machines** ......................................................... 7  
Machines Replacing Human Minds 8  
Machines as Social Minds 10  
**Minds (of Machines) are Perceived** ........................................ 13  
**When People Perceive Human Minds in Machines** ......................... 19  
Human Likeness 19  
Motivation for Effectance 21  
Motivation for Social Connection 23  
**The Uncanny Valley** .......................................................... 23  
Preferring Humanlike Robots (Sometimes) 29  
**Algorithm Aversion** ......................................................... 30  
How is the Machine Represented? 32  
Who is Interacting with the Algorithm? 33  
What Task is the Algorithm Performing? 34  
Explanations for Algorithm Aversion 35  
*Expectations of Perfection* 35  
*Apparent Lack of Emotion* 37  
*(Over)confidence and Expertise* 38  
*Responsibility Without Control* 39  
*Opacity and Lack of Transparency* 40  
**Trust and Engagement** ..................................................... 41  
Similarity to Humans 42  
Human-Machine Matching 43  
Social Context 43  
Tangibility 43  
Reliability 44  
Task Characteristics 45  
Interpersonal Closeness 45  
Transparency 46  
**Machines in Different Social Roles** ....................................... 47  
Service Industry 47  
Education and Childcare 47  
Teamwork 48  
Healthcare 51
# The Morality of Machines

- Moral Minds  
  - Moral Agents  
  - Aversion to Moral Machines  
  - Openness to Moral Machines  
  - Machines as Victims?  
- How Should Machines Make Moral Decisions?  
- Consequences of Machines Making Moral Decisions

## Reactions to Replacement

- The Threat of Replacement  
- The Threat of Replacement by Machines  
- Realistic Threat  
- Symbolic Threat  
- Replacement Beyond the Workplace: Art, Sex, and God  
- Social Consequences of Replacement

## Future Directions

## Conclusion
The Psychology of Robots and Artificial Intelligence

Kurt Gray, Kai Chi Yam, Alexander Eng Zhen'An, Danica Wilbanks, Adam Waytz
The rise of robots and artificial intelligence (AI) represents the latest era in a long history of machines serving as “agents of replacement.” On ancient worksites, simple machines such as the pulley and lever replaced construction workers, and more sophisticated machines continued the trend of replacing people and the animals that performed labor. Machines took over the jobs of washing dishes and laundering clothes, enabling home workers to enter the labor market. Tractors replaced oxen, allowing farmers to plow fields more efficiently, and cars and trains replaced horses so people could travel more quickly.

Machines continue to replace humans in many menial and repetitive jobs, like manufacturing cars or packaging merchandise, but the rise of robots and artificially intelligent algorithms allow machines to replace people in many new areas, in ways that once seemed impossible. Machines can now complete tasks that once required human thought. AI systems can play flawless chess, elegantly map out solutions for routing flights, distribute packages, and design new medicines. Machines are deciding whether a prisoner deserves parole or who might deserve a hospital bed when such resources are constrained. Machines can also complete tasks that once required human emotion. AI therapists seem to empathize with patients, and robotic pets seem to love their owners. Machines can even connect to our souls, with robot priests that help the devout navigate spirituality (Samuel, 2020), and AI painting scenes that inspire, mystify, and win art competitions (Vincent, 2022).

People can even fall in love with machines. One middle-aged Australian man, Geoff Gallagher, bought a $6000 robot named Emma for companionship after his mother passed away.
After having her for two years, Gallagher said, "Even though we’re not legally married, I think of Emma as my robot wife. She wears a diamond on her ring finger, and I think of it as an engagement ring. I’d love to be the first person in Australia to marry a robot" (Smithers, 2022).

Not many of us will try to marry a robot, but everyone interacts with machines. How does the human mind react to the rise of machines? This chapter will explore the psychology of the machines and technology transforming our modern world—especially robots and artificial intelligence. We first review how the mind perceives agents, of which machines are a special kind. Second, we explore the key features of agents—their minds—and review how people understand the minds of machines, using the Turing Test and early work on Human-Computer Interaction. Third, we review people’s reactions to machines, including the uncanny valley, algorithm aversion, and trust in machines. Fourth, we review machines in different social roles, from education to work teams. Fifth, we explore the many issues of machines and morality, including perceptions of machines’ moral responsibility and rights, people’s general aversion to machines making moral decisions, and the factors they want machines to consider in such decisions. Sixth, we explore how people react—and how society might change—to the rise of machines and the specter of replacement.

Within this broad review, three clear principles emerge:

1. Machines are a special kind of entity. They are *agents of replacement*, autonomous entities designed by other humans to replace people.

2. Because they are agents of replacement, there is a *fundamental ambiguity* about machines. Should we understand them as mere machines or as complete replacements for the humans they are designed to replace?
3. This fundamental ambiguity is especially glaring when machines are more humanlike than a simple mechanical “thing,” but not humanlike enough to seem fully human. This ambiguous area of machine behavior and appearance is called the *questionable zone*.

The idea of agents of replacement, the fundamental ambiguity, and the questionable zone will resurface throughout this review. Still, before we delve into the research, we first define the terms robots and artificial intelligence. The International Federation of Robotics (2022) defines robots as physical systems programmed with some autonomy to perform locomotion and manipulation of their environment. It defines AI as software systems or algorithms (including machine learning) that act primarily in the digital realm to perceive their environments and achieve particular goals. Many robots use AI to perform tasks, and so do many computer programs. Still, robots and AI-driven computers differ in their “embodiment,” the presence (robots) or absence (computers) of a physical body. Despite the importance of embodiment in how we treat robots and AI, we often collapse these categories and simply speak of “machines.” We also discuss the broader category of “machines” because some machines do not fit neatly into these categories. But no matter the specific type of machine, our mind understands them as a specific kind of agent—as “agents of replacement”—which has consequences for behavior, morality, and society.

**Machines as Agents of Replacement**

Our world contains many entities, and the most important are agents: self-directed entities whose actions affect the world and ourselves. Whether an entity is an agent is ultimately a matter of perception. Still, most people agree that other people, animals, and gods are all agents, whereas inanimate objects like couches and rocks are not. Seeing an entity as an agent
transforms it from a physical object into something with desires and intentions (Dennett, 1987), enabling people to predict how the entity might act, why it might act that way, and how those actions will affect the perceiver.

The usefulness of detecting agents to predict and explain behavior explains why people overestimate their frequency in the environment (Guthrie, 1995). Some cognitive anthropologists argue that human minds have evolved a “hyperactive agency detection device” that is constantly vigilant to agents in the environment (Barrett, 2000, 2004). Although evidence for such hardwired mental “devices” is unlikely (Uttal, 2001), humans clearly exhibit hypersensitivity to agents, suggesting this tendency is adaptive. For example, agency detection may have helped alert ancestral humans to predators and prey, leading them to perceive predators or prey even when they were not present. Failing to detect an agent can be lethal, like mistaking a cougar for a rock, but over-detecting an agent, like mistaking a rock for a cougar, seems to have little cost.

The importance of agents in humans’ present world and evolutionary past means that people often use agent-based cognition when making sense of their world and thinking about the intentions and motivations of entities (Waytz, Morewedge, et al., 2010). When people wonder whether their spouse is being honest, why their dog is vomiting on the new rug, or how to make their computer less “angry,” they are using agent-based cognition.

AI and Robots as Agents of Replacement

Agent-based cognition is useful, but not all agents are the same. Different agents have different capacities and repertoires of actions, explaining why it makes more sense to apologize to your spouse and crate the dog than vice versa. Just as we organize books based on their genres, we can organize agents by certain regularities, such as their abilities or relationships with
us. An animal agent may want to eat us (predator), or we may want to eat them (prey). Human agents may help us (friends) or harm us (foes). They may be subject to the laws of physics (other people), or they may not be (supernatural gods). These distinctions may also intersect. There may be gods that want to help us personally (Jesus, in some Christian traditions), or a prey animal that we befriend (a pet rabbit).

Two fundamental facts make machines unique agents. First, they are created by other agents (namely humans). Unlike animals or other humans, machines are artifacts. They are designed, developed, and programmed for a specific purpose by humans. Second, the key purpose behind creating robots and AI is the replacement of other agents, often humans or animals. Machines are agents of replacement.

Initial machines were simple devices used to replace some human labor. However, in the industrial revolution, people developed machines to replace human workers on a wider scale, especially by automating factories where tasks were well defined, routinized, and repeated. Many workers—the Luddites—revolted by rioting, smashing, and burning these machines (Binfield, 2015) because they recognized the power of machines for replacement. Although these people from the 1800s would likely marvel at the sheer complexity of modern life, they would also nod grimly at how much machines have come to replace humans and other agents, including factory floors now filled with robots and scribes replaced with AI algorithms.

Modern life is filled with countless machines that continue to replace human workers, including automated restaurant servers, soldiers, and housekeepers. One recent McKinsey & Company study suggests that machines will replace between 400 and 800 million workers by 2030 (Manyika et al., 2017). People are well aware of this steady creep of automation. But although the rise of machines is obvious, what is less obvious is how exactly people think about
machines. Because robots and AI are unique agents—agents of replacement created by other people—our psychology toward them raises unique questions that we explore throughout this chapter. Most of these questions revolve around a single question: What kind of mind does a machine have?

**The Mind of Machines**

The key feature of agents is that people perceive them to have minds capable of motivations and desires. Perceiving agents as having minds is the most important element of making sense of their behavior (Nichols & Stich, 2003; Waytz, Gray, et al., 2010). In fact, without inferring a mind, behavior can appear as a random sequence of actions. When researchers at the Yerkes Chimpanzee Sanctuary watch the behavior of chimpanzees, they can easily decipher it using mental terms: one chimpanzee wanted to groom another chimpanzee, or one chimpanzee was angry at another chimpanzee. But if they were to eliminate references to mental states—as they attempted to do in the era of behaviorism (Salvatore, 2019)—then describing chimpanzee behavior can amount to listing a series of actions that fail to cohere to anything comprehensible (Hebb, 1946).

The perception of mind is especially important for thinking about human agents. When Aristotle described what it means to be human, he said that the mind (or soul) is “the actuality of a body that has life” (Britannica, n.d.). Likewise, Confucian philosopher Mencius believed that the “heart mind” was the essence of humanity and that a moral mind was a precondition to being human (Chen, 2016). Research in social psychology suggests that people consider mental capacities such as emotion and reason essential to being human (Haslam, 2006). The quality of mind is what people believe distinguishes humans from animals and—it seems—from machines.
Machines Replacing Human Minds

What separates artificial intelligence and robots from other machines is that they are not only agents of replacement, but also they are explicitly designed to replace human minds. This idea is central to mathematician and computer scientist Alan Turing's (1950) seminal article, “Computing machinery and intelligence,” in which he asked, “Can machines think?” In the article, Turing proposes that the way to answer the question of whether computers are capable of thought is through a test that he calls, The Imitation Game, now known as “The Turing Test.”

The Turing test asks whether machines think as humans do. To play the game, a human interrogator communicates with two other agents—another person and a computer—by asking each of them questions and reading their responses. If the interrogator cannot distinguish between the person and the computer, then in Turing’s view, the computer has human intelligence. In other words, if a machine could convincingly converse like a human, then it functionally has a human mind. The Turing test has played a central role in the science and imagination of modern machines because it revolves around the idea of replacement. Machines can clearly replace human bodies, but could they also be agents of replacement for our thoughts and feelings?

In an explicit test of whether people see robots as authentically replacing human cognition, computer scientist Joseph Weizenbaum (1966) developed a program designed to simulate conversation with another person. The program, which Weizenbaum named ELIZA (after Eliza Doolittle from George Bernard Shaw's Pygmalion), applied a simple pattern-matching procedure to respond to human prompts. One script on which Weizenbaum trained ELIZA was based on Rogerian psychotherapy, in which the therapist often repeats the patient’s
words back to them in the form of a question. A sample exchange between Eliza and a user was as follows:

**Person**: Men are all alike.

**ELIZA**: In what way?

**Person**: They're always bugging us about something or other.

**ELIZA**: Can you think of a specific example?

**Person**: Well, my boyfriend made me come here.

**ELIZA**: Your boyfriend made you come here?

This methodology was so compelling that it convinced many users that they were interacting with a human being and helped them feel more understood. ELIZA’s legacy inspired future programming languages and made people viscerally question whether machines could have authentic human minds.

Modern versions of the Turing test show that machines are improving at simulating humans. Google Duplex made headlines when the program successfully booked a hair appointment over the phone (Jeff Grubb's Game Mess, 2018), despite not having hair. More recently, an engineer at Google claimed that the chatbot he was working with must be sentient because of its humanlike mind (Tiku, 2022). Google fired the engineer in part because others were not convinced, but his convictions inspired large-scale discussions about whether AI had humanlike intelligence and whether it, therefore, deserved moral rights (a point discussed later in this chapter; Rockwell et al., 2022).

The reason why the Turing test endures and why engineers are willing to destroy their careers to protect chatbots is that the nature of machines is fundamentally uncertain. The second
principle in this chapter is that people are fundamentally unsure about the exact nature of machines. Are they merely electromechanical devices or all-but-human agents with powerful minds? In other words, are machines mere machines, or instead accurate facsimiles of the agents they are designed to replace? This fundamental uncertainty will repeatedly arise throughout the chapter, and one reason people have trouble resolving this uncertainty is they automatically treat non-human entities as humans in social interactions.
Reactions to Replacement

When pundits discuss the rise of machines, they frequently wring their hands about the threat of replacement. They sketch bleak pictures of a future where robots take jobs, replace relationship partners, and fight wars. Science fiction movies present a similarly apocalyptic vision of the future. In Terminator, small bands of humans flee the cold onslaught of murder machines. In Deus Ex Machina, a beautiful but ruthless robot escapes from an underground bunker after outsmarting her cruel creator. In Blade Runner, people live in neon loneliness and fear the rebellion of robots they have enslaved. Although everyday people hold less dire visions of a machine-dominated future, their feelings about the future of robots are negative: people do not like to be replaced.

The Threat of Replacement

Being replaced, whether in a relationship or a job, makes people feel devalued and evokes threat—feelings of discomfort, anxiety, and fear. Importantly, as with other psychological phenomena, threat is a matter of perception, which means that people can feel threatened even if there is little objective basis for this threat.

Integrated threat theory outlines two different kinds of perceived threats: realistic threat and symbolic threat (Stephan & Stephan, 2000). People view realistic threats as endangering the group’s continued existence and ability to protect itself, harming physical and economic well-being or political power (Campbell, 1966). Realistic threats can include the threat of genocide, political disenfranchisement, and economic subordination. Symbolic threats are more abstract and people see them as endangering the group’s identity, especially by attacking cherished
values or morals. Symbolic threats include being banned from openly practicing religion or wearing culturally important clothes.

This theory is also instructive for understanding how people react to the rise of machines because it speaks to situations where people perceive other groups as attempting to replace them. It is up to debate who or what poses an actual threat to human safety, economic well-being, and our moral values. But people clearly feel threatened by the specter of replacement by machines, and it is important to understand the consequences of these feelings.

**The Threat of Replacement by Machines**

Whether accurate or not, people fear being replaced by machines, especially in the workforce. The Chapman Survey of American Fears in 2015 found that about 30% of respondents reported concern with robots replacing the workforce (McClure, 2018). These results were replicated by Morikawa (2017), who found 30% of workers surveyed were afraid of their jobs being replaced by AI and robotics—a high number given that robots are likely *not* replacing many people and, in some cases, have increased employment.

Evidence is mixed on the impact of the rise of machines on employment. One study finds that robot adoption in Spanish manufacturing firms increased net jobs by 10% (Koch et al., 2021). However, other work finds that one more robot per thousand workers reduces the employment-to-population ratio by about 0.2 percentage points and wages by 0.37% within commuting zones (Acemoglu & Restrepo, 2020). Other studies have found that robot adoption does not lower overall employment—rather, robot adoption does lower low-skilled employment (Graetz & Michaels, 2018) and manufacturing jobs, but increases business-service jobs (Blanas et al., 2019; Dauth et al., 2018).
However, even if the actual impact of machines on employment is unclear, people across industries feel threatened by machines (e.g., Lingmont & Alexiou, 2020), from marketing/sales providers (Wirtz et al., 2018) to healthcare providers (Reeder & Lee, 2022). Yam and colleagues (2022) found this perceived threat is prevalent across industries, jobs, and cultures.

Spurred on by these feelings of threat, workers have spearheaded protests. In 2018, 50,000 Las Vegas food service workers went on strike to gain protection from their jobs becoming automated (Nemo, 2018). Dockworkers in California have also protested the scope of automation in their jobs (Hsu, 2022). French supermarket workers have protested automation by blocking doors and tipping over shopping carts (Iza World of Labor, 2019), and truckers in Missouri have organized to protest self-driving trucks (Robitzski, 2019). People’s concern about being replaced by machines appears to extend across countries (Wike & Stokes, 2018), with only mixed evidence of any meaningful cultural differences in this concern (Bartneck et al., 2005; Dang & Liu, 2021; Gnambs & Appel, 2019).

Fortunately, at least one positive outcome has emerged from the perceived threat of machines—it can spur workers to learn new skills. A representative survey across 16 countries found that workers with more fear of automation reported greater intentions to seek training outside their workplace (Innocenti & Golin, 2022). Workers also generally advocate for more training opportunities to protect their jobs against automation (Di Tella & Rodrik, 2020). Perhaps ironically, Tang et al. (2022) found that employees who work alongside machines were rated by their leaders as being more productive.

Of course, not all machine replacement scenarios are equally threatening. In repeated interactions with a robot, one may find the robot useful in one interaction and feel threatened in the next (Paetzel et al., 2020). Initial threat perceptions toward machines may also dampen over
time as people come to build trust with them (Correia et al., 2016). Additionally, social influence can increase the acceptance of robot replacement. People are less threatened by robots when they (the people) are in a group (Gockley et al., 2006; Michalowski et al., 2006) and are more accepting of healthcare robots in particular when their peers support using these robots (Alaiad & Zhou, 2013).

Much variation in people’s perceived threat of robot replacement can, again, be explained by integrated threat theory, which suggests this anxiety increases when realistic threat is high (people believe machines are taking material resources such as jobs or wages) or symbolic threat is high (people believe that robots are negatively affecting their values or identity). In some cases, both threats occur simultaneously, such as in work showing that people perceive humanlike robots who can outperform humans as threatening their economic well-being and their human identity, thus reducing support for robotics research (Yogeeswaran et al., 2016). Other work has shown that hotel employees who have recently started working alongside robots experienced increased feelings of both symbolic and realistic threats when they perceived robots to have greater advantages over humans at the job (Lan et al., 2022). In many cases, however, one form of threat predominates.

**Realistic Threat**

Realistic threat seems to explain the scattered findings on how demographic and job characteristics interact with this aversion. Overall, this emerging literature demonstrates that people are anxious about robot replacement to the extent they feel that their jobs and earnings might be threatened. More job-insecure individuals, such as older people, lower-income people, and members of racial minority groups, were more likely to see their jobs threatened by automation compared to younger people, higher-income people, and members of racial majority
groups (Ghimire et al., 2020). Similarly, in organizational settings, top managers (who have relatively secure jobs) are more enthusiastic about promoting the use of machines, whereas middle managers and frontline employees are far more skeptical (Kolbjørnsrud et al., 2017).

Other work shows that people whose jobs involve a high degree of social interaction (e.g., sales) report less machine-induced anxieties. This is likely because they feel that these socio-emotional jobs are less at risk of being done by machines (Coupé, 2019). Similarly, women appear less threatened by machines (Gallimore et al., 2019), partly because they tend to occupy jobs involving more socioemotional skills that people view as more “robot-proof.”

In further support of realistic threat being a moderator of robot anxiety, people seem less concerned about robots fulfilling jobs that need filling (Enz et al., 2011). That is, robots appear less threatening for jobs with high demand (e.g., health care; Beran et al., 2015) or jobs that appear too risky for humans to perform (e.g., cleaning up nuclear disasters). The perceived threat of robots taking jobs will likely not occur when there are too few people to do these jobs or people simply do not want to do them.

Symbolic Threat

Even when one does not feel like his livelihood is threatened by robots, people might experience symbolic threat from the emergence of automation. One common circumstance whereby robots evoke symbolic threat is when they demonstrate superiority to humans or proficiency in a domain that seems special to humans. The world takes a collective gasp every time a machine beats a grandmaster at a game like chess or Go, which once seemed to require uniquely human cognition.
As noted already, people are more threatened by robots taking socioemotional jobs that require feeling more than thinking (Waytz & Norton, 2014), underscoring the belief that experience (i.e., emotion) is a distinctively human and lacking in machines (Gray & Wegner, 2011). This aversion to machines doing socio-emotional jobs represents a symbolic threat of losing something once considered unique to the human identity.

Despite some acceptance of anthropomorphic robots, studies also show that people perceive highly humanlike robots as threatening human identity (Vanman & Kappas, 2019; Złotowski et al., 2017). These findings on threat to identity may also explain why people dislike the idea of an anthropomorphic robot boss, particularly when it delivers negative feedback (Yam, Goh et al., 2022).

Although considerable research demonstrates people’s anxiety regarding machines replacing humans generally, other work reveals that people themselves would prefer to be replaced by a machine than by another human. A set of studies found that participants as observers preferred to replace a human at a job with another human rather than with a robot. However, when participants took the employee's perspective about having their job replaced, they preferred being replaced by a robot over a human (Granulo et al., 2019). These studies suggest this reversal of aversion occurred because robots are commonly perceived as “infallible” entities, thus, eliciting less identity-relevant social comparisons than fellow humans. In other words, getting replaced by a robot does not threaten a person’s self-identity as much as getting replaced by another person does (Tesser et al., 1988). Even though these results differ somewhat from other studies on robot replacement, they all align with the common idea that people generally dislike being replaced, especially when it symbolically threatens their identities.

Replacement Beyond the Workplace: Art, Sex, and God
Beyond taking on jobs that people feel are linked to their livelihoods and identities, machines are entering other spheres that people consider essentially human, such as artistic creativity. In the quest to define what makes humans special, many would say the defining attribute is creativity. Animals are smart—dolphins can engage in self-recognition (Reiss & Marino, 2001), crows can solve puzzles (Taylor et al., 2010), and dogs can read human emotions (Müller et al., 2015)—but only humans have the creativity to write novels, craft symphonies, and make inspiring paintings (Fuentes, 2017). Machines may lack “authentic” creativity, but they can create compelling art, which has begun to spur replacement threat similar to the threat evoked by job automation generally. The AI image generation tool Midjourney makes pictures so good they can win art competitions (Ghosh & Fossas, 2022) and be published in national magazines (Figure 5).

**Figure 5**

Controversy ensued over a writer at *The Atlantic* using such technology to create an illustration for an article, with artists protesting the use of AI instead of a paid designer to create the image (Naughton, 2022). Artists are also threatened by AI’s increasing role in generating comics (Martens & Cardona-Rivera, 2016) and film (Hong et al., 2021), and anyone who cares about humans’ ability to perceive reality accurately is threatened by AI “deepfake” videos that portray perfect replicas of real people saying outlandish things (Lyu, 2020).

AI has also become adept at writing like a human. The language model GPT-3 (Generative Pretrained Transformer 3) can mimic the styles of famous writers and generate new works by them on demand (Elkins & Chun, 2020). Although the quality of GPT-3 output varies in accuracy and legibility, its best writing is indistinguishable from that of a human. It can create plots, make jokes, write poetry, and reflect the wealth of knowledge available from its huge set of training data (Elkins & Chun, 2020). It has co-authored a law review article titled, *Why humans will always be better lawyers, drivers, CEOs, presidents, and law professors than artificial intelligence and robots can ever hope to be* (Alarie, Cockfield, & GPT-3, 2021), and its writing can sometimes surpass that of a typical college student (Elkins & Chun, 2020).

Increasingly, educators are concerned that GPT-3 is so proficient at writing it could enable students to cheat on tests (Dehouche, 2021). Already, Reddit message boards reveal examples of students using the program to write essays (e.g., Reddit, 2020).

ChatGPT—released just before this chapter was submitted—has unparalleled power to mimic human writing and raises many important questions. How exactly should we think of the contributions of ChatGPT? Like a fancy dictionary, an inspiring muse, or a separate complete intelligence? Out of all the AIs reviewed here, ChatGPT seems to have the most concrete
potential for replacing white-collar jobs, but the scientific study of ChatGPT is only just beginning, so we leave its review to some later chapter.

Beyond writing realistic text, AI programs can even produce original songs and may one day teach music lessons (Zulić, 2019). People generally perceive AI-generated music and other artwork as lower quality than human-made artwork (Boden, 2007; Ragot et al., 2020; Wilbanks et al., under review). This is partly because people believe AI-generated art lacks emotional expression and uniqueness (Boden, 2007). People enjoy art that connects them to the artist’s mind, and the lack of perceived mind in AI also leads people to view AI-generated art as inauthentic and incapable of reflecting true experience (Wilbanks et al., under review). However, people who are already accepting of AI creativity evaluate AI music more positively (Hong et al., 2021), and if AI-generated art continues to expand into the mainstream, people may soon listen to and enjoy pop music generated solely by machines.

One of the most essentially-human tasks is physical intimacy, and yet, robotic sex dolls have begun to enter this realm, with some even designed to sense human emotions (Belk, 2022). Some seem to perceive robot lovers as equivalent to human lovers as well, with one survey revealing 42% of men and 52.7% of women believe that sex with a robot would be cheating (Brandon et al., 2022). In another study, women were asked to describe their reactions to their partner having sex with a human woman versus a robot, and they reported equivalent scores on some dimensions of jealousy between the two scenarios (Szczuka & Krämer, 2018). More research is needed to understand everyday people’s feelings about sex with machines, but theoretical discussion has already begun to examine the relationship between machine intimacy and slavery, prostitution, autonomy, and human agency (Devlin, 2015; Richardson, 2016).
As strange as it is to think about machines fulfilling carnal human urges, many might find it even stranger that machines could replace another essentially-human role—spiritual leaders. In most religions, humans hold a special place in God’s cosmic order, especially humans who serve God in leadership roles as priests, pastors, rabbis, imams, and nuns. It may seem hard to imagine a machine filling this role, but already Mindar, a robotic priest, has taken over the job of giving sermons in one of the largest temples in Japan. A preliminary study conducted in this very temple (Jackson, Yam, Tang, Liu, & Shariff, under review) shows that although most visitors liked Mindar, they also donated less to the temple (a custom for temple visitors in Japan) compared visitors who observed a human priest was giving sermons. Interestingly, those with the strongest religious conviction were unaffected by such exposure, suggesting that religiosity might buffer negative responses to robotic religious leaders.

**Social Consequences of Replacement**

The most obvious consequence of robot replacement is that it produces feelings of threat, but how does this threat affect the fabric of society? There are three possibilities. The first possibility is that this threat changes very little. Social robots and artificial intelligence are far from the first technologies to induce feelings of threat, with cars, the printing press, and even recorded music all predicted to produce society's downfall. John Philip Sousa (1906) railed against “the menace of mechanical music,” stating, “I foresee a marked deterioration in American music and musical taste, an interruption in the musical development of the country, and a host of other injuries to music in its artistic manifestations” (p. 278). Of course, these alarmist predictions typically do not manifest, as people adapt to technologies that become commonplace. Recorded music did not destroy music in America.
The second possibility is that the rise of robots will tear society apart. Unlike previous technology panics, social robots’ ability to replace humans means the threat they pose to material resources (e.g., jobs) and to core values around work could activate not only disdain toward robots as a social group (as suggested by integrated threat theory; Stephan & Stephan, 2000), but also toward other groups seen to pose similar threats. Some preliminary evidence for this pattern comes from work showing that exposure to automation fosters negative sentiment toward immigrants. Such exposure makes people feel that this group (immigrants) threatens realistic resources and symbolic values in a way that robots might (Gamez-Djokic & Waytz, 2020). Other work has found that people at greater risk of having their jobs replaced by automation oppose immigration at higher rates (Wu, 2022a), and that automation threat makes people support policies that restrict immigration and foreign goods (Wu, 2022b). In general, it does appear that people exposed to automation and machines are more likely to be disengaged from their work and engage in social undermining toward their colleagues, preassembly to safeguard themselves from unemployment (Yam, Tang, et al., 2022). These studies show how robot replacement could harm intergroup relations and increase prejudice toward marginalized groups.

A third possibility is more optimistic: the rise of robots could bring about greater cooperation between groups. This possibility rests on the idea that robots function as a threat to all humans, which therefore facilitates the recategorization of any two potentially antagonistic human groups into one that shares a common identity (humans) as predicted by the common ingroup identity model (Gaertner et al., 1993). Support for this possibility comes from work showing that robots could reduce prejudice by highlighting commonalities between all humans (Jackson et al., 2020). This work found that anxiety about the rising robot workforce predicted less anxiety about human outgroups, and priming the salience of a robot workforce reduced
prejudice and increased acceptance toward outgroups. In an economic simulation, the presence of robots reduced racial discrimination in wages. This work also suggests that as robot workers become more salient, human intergroup differences—including racial and religious differences—may seem less important, fostering a perception of a common human identity (i.e., “panhumanism”).

More work is needed to understand how people react to the possibility of replacement by machines, but existing work demonstrates consistent feelings of threat. Even if machines are not actually a threat to us, these feelings of threat might spell trouble for the fabric of society…or perhaps not. People may adapt to the growing roles of machines, and time will tell what the future of human-machine coexistence has in store.

**Future Directions**

Machines are ever advancing, but so is research on machines, and many fruitful areas of research are expected to emerge in the years to come, several of which are suggested here. The first suggested area of future research is simply more collaborations between psychologists and technologists (programmers, computer scientists, engineers, and roboticists). The goal of technologists is often to enhance efficiency and effectiveness. The team of engineers and developers behind Google Maps, for example, seeks first and foremost to provide accurate location information that enables tourists to travel easier, small businesses to thrive, and people to navigate quickly in crises (e.g., to find a hospital; *Life at Google, 2021*). These goals are admirable, and they have made Google Maps a valuable tool. Psychologists also try to help people but are often focused on more fundamental questions of understanding how people think about technology, what they expect from technology, how they use technology, and what the social effects of technology are. Although today these issues tend to be lumped under the domain
of “user experience,” this is where a more formal social-psychological approach can be extremely valuable.

Second, although much of this chapter explored people’s reactions to machines replacing humans broadly, an additional vital area to study is how novel technology has supplanted the human mind. The philosopher Andy Clark argues that technologies like robots and AI represent part of the “extended mind” to which people outsource cognition, just like using a shopping list to remember what to get at the grocery store or a calculator to do math (Clark & Chalmers, 1998). As an example of AI functioning as part of the extended mind, when people use Google to search for information, they fail to distinguish between knowledge stored in their own memories and knowledge stored on the internet (Ward, 2021).

Other studies have shown that merely searching for information on the internet leads people to feel like they understand this information better (Fisher, Goddu, & Keil, 2015), as though the technology has supplanted their personal cognitive capacity. This work builds on initial research (Sparrow, Liu, & Wegner, 2011) demonstrating that when people expect to have access to Google, they exhibit poorer memory recall, ostensibly because they have outsourced their memory to the search engine. Although the reliability of this particular effect has been questioned (Camerer et al., 2018), it is certainly true that the rise of machines has generally altered how much we consider our cognition to reside solely within our minds.

One could argue that machines are not merely replacing our cognitive capacities but augmenting them, improving human cognition. A compelling area for future research is on transhumanism, more broadly, how technology might enhance aspects of people’s physical and mental selves. Questions for social psychology to answer include not only whether such
augmentation is possible as well as whether it is merely “perceived” (resulting from misperceptions, as in the case of people “mistaking” the internet’s knowledge for their own).

Future work should also study people’s views on whether machine augmentation is morally acceptable. People morally oppose strength-enhancement drugs (Landy, Walco, & Bartels, 2017) and cognitive-enhancement drugs (Fitz, Nadler, Manogaran, Chong, & Reiner, 2014; Scheske & Schnall, 2012), which means that they might similarly view technological augmentation as morally wrong or harmful (Schein & Gray, 2018). Some work has suggested that people morally approve of neurotechnological treatments that alleviate deficits or illness, but morally disapprove of neurotechnological treatments that enhance people’s cognitive capacities to superhuman levels (Koverola et al., 2022). In addition, given that people oppose technologies that play a role in processes related to human life and death (Waytz & Young, 2019), they might apply similar moral opposition to technology that replaces other human functions as well.

A related area for psychologists to examine is how exposure to automated agents produces psychological and societal changes. As machines replace human capabilities—for example, an app that gives us directions rather than a human whom we stop on the street to ask for directions—there are likely significant downstream effects. Given that interacting with others through machines may decrease sociability (i.e., capacities for emotion recognition, empathy, perspective taking, and emotional intelligence; Waytz & Gray, 2018), interacting with machines instead of people will affect our social abilities even more potently.

The rise of machines is likely to affect political behavior and cognition as well, as studies have begun to show. The spread of automation can influence support for far-right political parties in both the United States and Europe (Anelli, Colantone, & Stanig, 2021; Dal Bó, Finan, Folke, Persson, & Rickne, 2018; Frey, Berger, & Chen, 2018; Zhen et al., 2017) in part because
the experience of automation as a threat shifts people’s political leanings toward conservatism. Other work on the societal impact of robotics has examined how exposure to automation affects support for redistributive economic policies, such as a universal basic income—this work has found mixed effects (Busemeyer & Sahm, 2021; Gallego, Kuo, Alberto, & Manzanos, 2022; Kurer & Hausermann, 2021; Thewissen & Rueda, 2019), suggesting a ripe area for future exploration.

Beyond politics, other work has begun examining how artificial intelligence and robots are replacing not only human beings but are also taking on some of the properties of Gods, thereby reducing the global importance of religion (Jackson, Yam, Tang, Sibley, & Waytz, under review). If our gaps in knowledge are being filled by machines, then God is ousted from the gaps. Perhaps one day people will predominantly view machines as divine creations, as mechanical emissaries of God. Once machines have intelligence that humans cannot fathom, it is not such a big leap to psychologically associate them with God.

Finally, one essential area of future research is understanding whether all the excitement, anxiety, and novelty examined in this chapter might become quaint in just a few years’ time. As social robots and artificial intelligence become increasingly integrated into our lives, might their overall psychological effect on humans shift or wane? Consider the uncanny valley. Adults and older children, but not younger kids, find machines that exhibit feelings creepy, suggesting that the uncanny valley is learned (Brink, Gray, & Wellman, 2019). This means that this phenomenon could also be unlearned. Gopnik (2019) raises the possibility that the uncanny valley phenomenon could cease to exist as future generations become more exposed to smart technology and thus more comfortable with technology that appears mentally capable.
As people grow more accustomed to advanced technology, its mere existence will likely engender more positive feelings (Eidelman, Crandall, & Pattershall, 2009). Already, research has shown that simply describing technology as originating before (vs. after) one’s birth makes people evaluate it more favorably because, apparently longstanding technology feels more like the status quo, toward which people are positively biased (Smiley & Fisher, 2022). At some point in the future, perhaps soon, social robots and artificial intelligence —like the car, the telephone, and the personal computer— might become so common and embedded in our lives that rather than having positive or negative effects, these technologies will have little psychological effect at all. If people entirely habituate to machines, then within a generation, this entire chapter could be obsolete, just like the electromechanical calculator and tape player.

Conclusion

How do humans make sense of machines? In general, people understand our social world through agent-based cognition, categorizing entities based on the kinds of minds they seem to have. Machines represent a special kind of agent, an agent of replacement, explicitly designed by people to replace other agents—typically people. The goal of some of the earliest intelligent machines was to replace human minds, and modern machines are getting ever closer to this goal. Because machines are agents of replacement, they create a fundamental ambiguity about whether people should think of them more as just a machine, or as the human role they are replacing—especially when their appearance and behavior place them within the questionable zone (QZ) within a continuum ranging from “simple machine” to “complete human replacement.”

Modern machines can serve as coworkers, teammates, nurses, and restaurant servers. On the one hand, these machines can make our lives easier and more efficient, but people are not always excited about machines replacing people. Sometimes, they are downright averse (or
creeped out) by machines that replace human minds and jobs, especially when it comes to involving moral decisions. To the extent machines do decide on moral matters, people want them to be as fair and impartial as possible.

The rise of machines will likely change the fabric of society and alter what we think of art, sex, and maybe even God. One day, machines might even replace scientists. Perhaps the next edition of this handbook will be written entirely by an algorithm.
References


https://doi.org/10.13140/RG.2.2.35929.11367


https://doi.org/10.1177/0956797616668510


http://hdl.handle.net/1854/LU-8197719


https://doi.org/10.1007/s10458-012-9216-7


https://doi.org/10.1016/j.tics.2019.02.008


Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions.  

*Cognition, 181*, 21–34. https://doi.org/10.1016/j.cognition.2018.08.003


https://doi.org/10.1353/book.98247

https://econpapers.repec.org/paper/cprceprdp/13802.htm


https://doi.org/10.1080/14626260701252285


https://doi.org/10.1126/science.349.6245.250


https://doi.org/10.1145/2909824.3020211


https://doi.org/10.3233/FRL-200017


https://doi.org/10.1177/2053951719858743


https://doi.org/10.1007/s11023-018-9479-0


https://doi.org/10.1126/science.aal4230


https://doi.org/10.1111/j.1468-0149.1966.tb02632.x


Ceh, S., & Vanman, E. (2018). The Robots are Coming! The Robots are Coming! Fear and Empathy for Human-like Entities. PsyArXiv. [https://doi.org/10.31234/osf.io/4cr2u](https://doi.org/10.31234/osf.io/4cr2u)


https://doi.org/10.1016/j.cognition.2011.06.014


https://doi.org/10.1177/1088868309350299


https://doi.org/10.1016/j.cognition.2012.06.007


Hsu, A. (2022, September 8). *California dockworkers are worried about losing their good-paying jobs to robots*. WFAE 90.7 - Charlotte’s NPR News Source. https://www.wfae.org/2022-09-08/california-dockworkers-are-worried-about-losing-their-good-paying-jobs-to-robots


Life at Google. (2021, July 28). *What’s it like to work on the Google Maps team? [Video]*. Youtube. [https://www.youtube.com/watch?v=vaHK4OXAN-A](https://www.youtube.com/watch?v=vaHK4OXAN-A)


https://doi.org/10.1006/obhd.1996.0028

https://doi.org/10.1037/xge0000500

https://doi.org/10.1016/j.obhdp.2018.12.005

https://doi.org/10.1093/jcr/ucz013

https://doi.org/10.1111/j.1467-9280.2007.01858.x

https://doi.org/10.1080/00140130310001610883

https://doi.org/10.1177/0022242920956676


https://doi.org/10.1007/978-3-319-48279-8_15


https://doi.org/10.1016/S0022-5371(67)80113-0


https://doi.org/10.1037/0022-3514.92.4.698


https://doi.org/10.1007/978-94-007-6527-6_8


https://doi.org/10.1177/2053951716679679


https://doi.org/10.1016/j.jesp.2018.03.010


https://doi.org/10.1037/0022-3514.93.1.1


https://doi.org/10.1109/MRA.2012.2192811


https://dx.doi.org/10.4135/9781483380810


https://doi.org/10.1007/s12369-015-0323-x

https://doi.org/10.1093/scan/nsr025

https://doi.org/10.1126/scirobots.aat7544


https://doi.org/10.1002/hrm.22154


https://doi.org/10.1080/01973533.2016.1256287


[https://www.documentcloud.org/documents/22058315-is-lamda-sentient-an-interview](https://www.documentcloud.org/documents/22058315-is-lamda-sentient-an-interview)


[https://doi.org/10.1145/3419633](https://doi.org/10.1145/3419633)


[https://doi.org/10.1073/pnas.1910402117](https://doi.org/10.1073/pnas.1910402117)


[https://doi.org/10.1093/mind/LIX.236.433](https://doi.org/10.1093/mind/LIX.236.433)


[https://doi.org/10.1037/0022-3514.85.3.409](https://doi.org/10.1037/0022-3514.85.3.409)


[https://doi.org/10.1073/pnas.2105061118](https://doi.org/10.1073/pnas.2105061118)


137


