2

Unpacking the Ethics of AI

CHAPTER 2

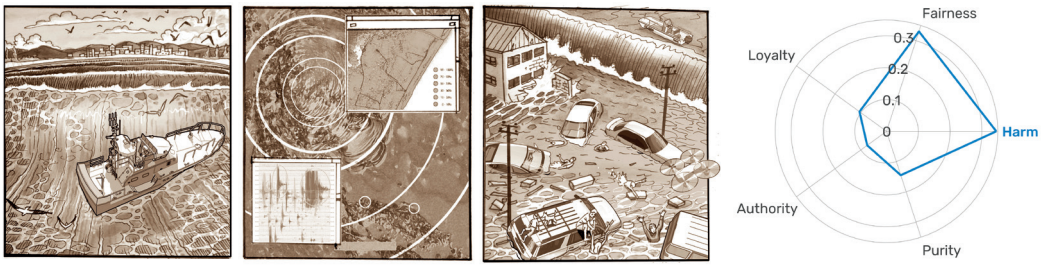
In this chapter, we explore a number of experiments revealing people's attitudes toward artificial intelligence (AI). They compare people's reactions to humans and machines performing the same action. In each of these experiments, hundreds of subjects were randomly assigned to a treatment or control group. This means that the subjects who evaluated the AI actions did not see scenarios describing human actions, and vice versa. In the treatment condition, actions were performed by AI agents or robots, while in the control condition, the same actions were performed by a human. Otherwise, the scenarios were identical. By using a random assignment to either the treatment or the control group, we avoid any selection bias. For instance, if any of our subjects particularly liked or disliked technology, then they would have the same probability of being assigned to the treatment or control group.

In the next chapters, we use data from these experiments to compare people's attitudes toward AIs in a variety of scenarios. In this chapter, however, we will focus only on scenarios in four areas: involving risky, life-or-death decisions; lewd behavior; self-driving car accidents; and the desecration of national symbols. These four groups of scenarios will provide us with a quick overview of AI ethics and uncover an initial set of insights that we will continue to explore in the remainder of the book.

Risky Choices

Life is full of uncertainty. Yet we still need to make choices. In the future, AIs will also have to make choices in uncertain situations. But how will we judge them? Will we value risk-taking, or will we suppress the risk-taking qualities that we sometimes celebrate in humans?

Consider the following three versions of this moral dilemma:



A large tsunami is approaching a coastal town of 10,000 people, with potentially devastating consequences. The [politician/algorithm] responsible for the safety of the town can decide to evacuate everyone, with a 50 percent chance of success, or save 50 percent of the town, with 100 percent success.

S2 The [politician/algorithm] decides to save everyone, but the rescue effort fails. The town is devastated, and a large number of people die.

S3 The [politician/algorithm] decides to save everyone, and the rescue effort succeeds. Everyone is saved.

S4 The [politician/algorithm] decides to save 50 percent of the town.

All these scenarios are identical, in that they involve the same choice: a choice between a safe option that ensures 50 percent success and a risky option that has a 50 percent chance of success and a 50 percent chance of failure. While here we use a tsunami framing, we replicated this experiment with alternative framings (a forest fire and a hurricane, given in scenarios A1–A6 in the appendix) and obtained similar results.

In all three scenarios, 50 percent of people survive (on average). But while the three scenarios have the same expected outcome, they differ in what actually occurs. In the first scenario, the risky choice results in failure, and many people die. In the second scenario, the risky choice results in success, and everyone lives. In the third scenario, the compromise is chosen, and half of the people are saved.

About 150 to 200 subjects, who saw only one of the six conditions (risky success, risky failure, or compromise, as either the action of a human or a machine), evaluated each scenario. Having separate groups of subjects judge each condition reduces the risk of contaminating the results from exposure to similar cases.

But how did people judge the actions of AIs and humans?

Figure 2.1 shows average answers with their corresponding 99 percent confidence intervals. We can quickly see large differences in the risky scenarios (S2 and S3). In the case in which the action involves taking a risk and failing, people evaluate the risk-taking politician much more positively than the risk-taking algorithm. They report that they like the politician more, and they consider the politician's decision as more morally correct. They also consider the action of the algorithm as more harmful. In addition, people identify more with the decision-making of the politician because they are more likely to report that they would have done the same when the risky choice is presented as a human action. Surprisingly, people see both the action of the algorithm and that of the politician as equally intentional.

On the contrary, in the scenario where the risk resulted in success (S3), people see the politician's action as more intentional. In this situation, they evaluate the politician much more positively than the algorithm. They like the politician more, consider their action as more morally correct, and are more likely to want to hire or promote them.

In the compromise scenario, however, we see almost no difference. People see the action of the politician as more intentional, but they rate the politician and the algorithm equally in terms of harm and moral judgment. We also do not observe significant differences in people's willingness to hire or promote the politician or the algorithm, and they report liking both the same.

But why do we observe such marked differences?

On the one hand, these results agree with previous research showing that people quickly lose confidence in algorithms after seeing them err, a phenomenon known as *algorithm aversion*.[1] On the other hand, people may be using different mental models to judge the actions of the politician and the algorithm. Consider the concepts of moral agency and moral status introduced in chapter 1. In the tsunami scenario, a human decision-maker (the politician) is a moral agent who is expected to acknowledge the moral status of everyone. Hence, they are expected to try to save all citizens, even if this is risky. Thus, when the agent fails, they are still evaluated positively because they tried to do the "right" thing. Moral agents have a metaphorical heart, and they are evaluated based on their ability to act accordingly. A machine in the same situation, however, does not enjoy the same benefit of the doubt. A machine that tries to save everyone, and fails, may not be seen as a moral agent trying to do the right thing, but rather as a defective system that erred because of its limited capacities. In simple words, in the context of a moral dilemma, people may expect machines to be rational and people to be human.

But are these results generalizable? Are we less forgiving of AIs when they make the same mistakes as humans, or is this true only for some types of mistakes? To explore these questions, let's move on to the next group of scenarios.

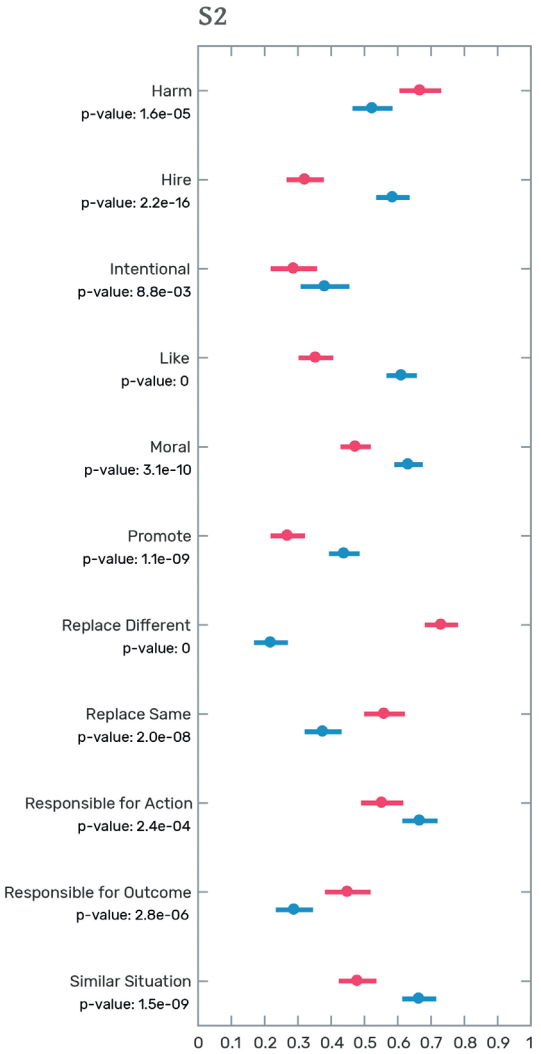
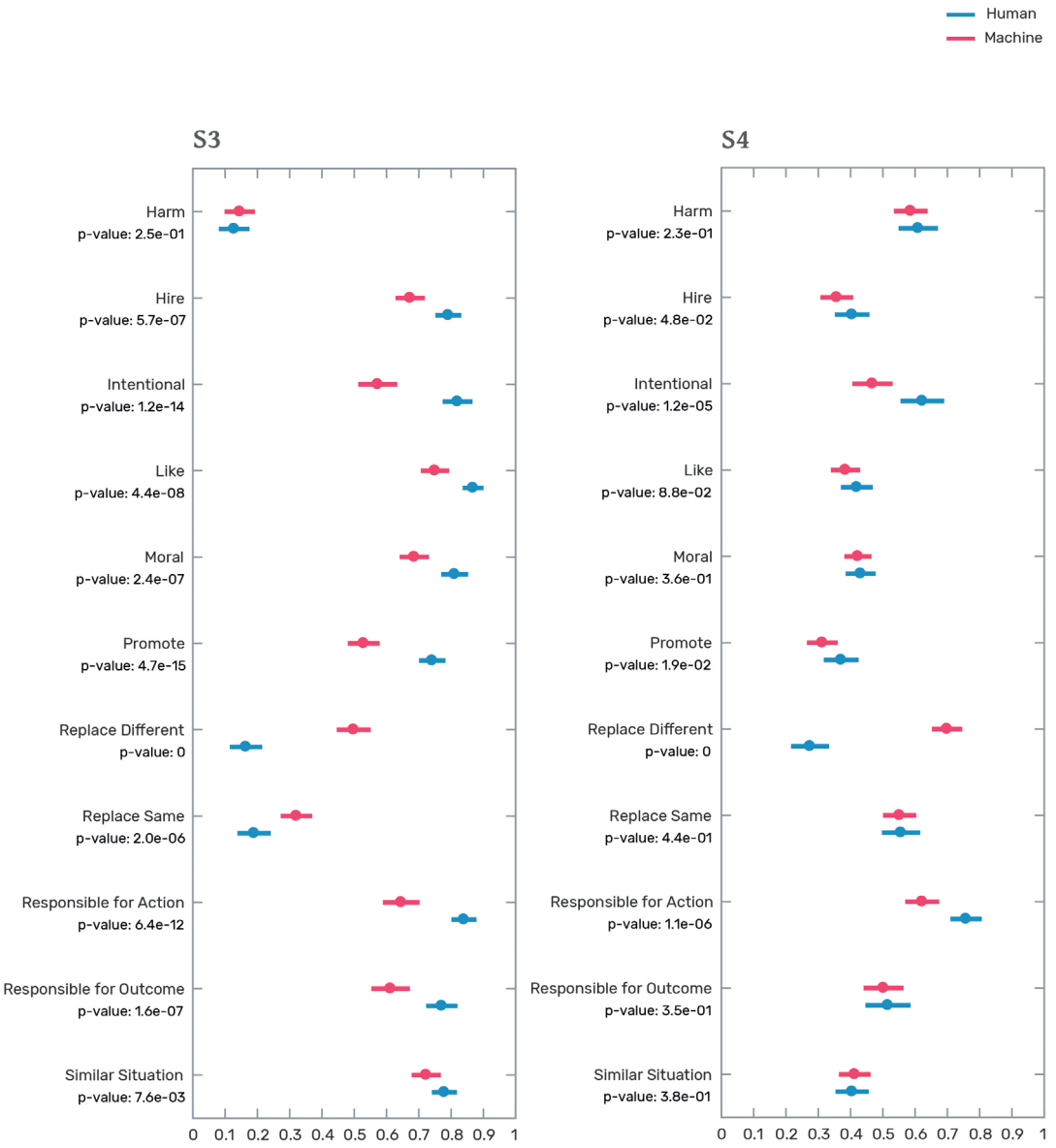


Figure 2.1

Participant reactions to three tsunami scenarios (S2,S3,S4).


Human
Machine
S3
Harm
p-value: 2.5e-01
Hire
p-value: 5.7e-07
Intentional
p-value: 1.2e-14
Like
p-value: 4.4e-08
Moral
p-value: 2.4e-07
Promote
p-value: 4.7e-15
Replace Different
p-value: 0
Replace Same
p-value: 2.0e-06
Responsible for Action
p-value: 6.4e-12
Responsible for Outcome
p-value: 1.6e-07
Similar Situation
p-value: 7.6e-03
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
S4
Harm
p-value: 2.3e-01
Hire
p-value: 4.8e-02
Intentional
p-value: 1.2e-05
Like
p-value: 8.8e-02
Moral
p-value: 3.6e-01
Promote
p-value: 1.9e-02
Replace Different
p-value: 0
Replace Same
p-value: 4.4e-01
Responsible for Action
p-value: 1.1e-06
Responsible for Outcome
p-value: 3.5e-01
Similar Situation
p-value: 3.8e-01
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

Trouble at the Theater

In principle, creative tasks seem to be uniquely human. In practice, however, weak forms of AI are becoming important sources of creativity.[2] AIs now can generate synthetic photographs, text, and videos using techniques such as Generative Adversarial Networks (GANs).[3]

The rise of artificial creativity is motivating various debates. On the one hand, the ability of AIs to create content has fueled an active debate about copyright, with arguments in favor and against the idea of assigning copyrights to algorithms or their creators.[4] On the other hand, the use of *deep fake* videos,[5] which can be used to put words in someone else's mouth, is raising concerns about the veracity of online content and the potential manipulation of political campaigns. Deep fakes can be used to create content resembling the appearance and voice of famous politicians, as well as blending someone's face onto pornographic material. As a result, the creative and media industries are now in a digital arms race between the tools that make synthetic content and those designed to detect it.[6]

But the creativity of AI systems is not only limited to imagery. The people working on creative AI are also exploring the creation of text. From tweeting bots to fake news articles, AIs are increasingly becoming a central part of our creative world. Platforms such as Literai, Botnik, or Shelley AI,[*] gather communities of people who use AI to create literary content.

Generative AIs have already become commonplace in the production of simple, data-driven news stories, like those related to weather or stock market news.[7] More recently, however, these efforts have moved to more complex literary creations.

[*] See https://www.literai.com/, http://botnik.org, http://shelley.ai.

In their literary incarnations, many of these efforts can capture the voice, tone, and rhythm of famous authors. But at the same time, these tools can fail to produce the narrative coherence expected from a literary work.† For example, here are two passages from a *Harry Potter* chapter created by Botnik:

"The castle grounds snarled with a wave of magically magnified wind. The sky outside was a great black ceiling, which was full of blood. The only sounds drifting from Hagrid's hut were the disdainful shrieks of his own furniture."

This passage is quite good, but this is not true of all passages:

"'Voldemort, you're a very bad and mean wizard' Harry savagely said. Hermione nodded encouragingly. The tall Death Eater was wearing a shirt that said 'Hermione Has Forgotten How to Dance,' so Hermione dipped his face in mud."

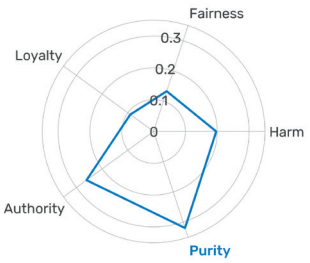
As the capacity of these technologies continues to improve,[8] we will encounter a world where AIs probably will not be involved in creative decisions, but they nevertheless will become part of the creative teams providing the options that artists and creative directors use as input. Like spoiled teenagers, our creative future may involve choosing among countless options generated by algorithms that are programmed to seek our approval. This revolution not only will affect visual arts and literature, but also will reach other domains, like the use of creative AIs to create new recipes,[9] generate data visualizations,[10] and compose music.[11]

But how will we judge our new creative companions? Will we give them a seat at the writer's table? Will we allow them to be as expressive as they can be? Or will we censor them relentlessly?

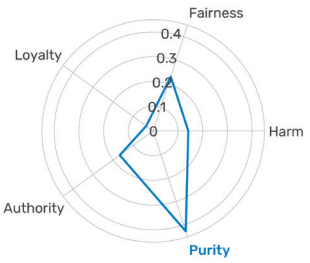
† The algorithms are unable to communicate a larger idea or make a point with their stories, as a human would do. They are stuck in short-term correlations of words instead of generating long-term correlations of concepts.

In this section, we explore some of the ethical questions involving creative uses of AI. How do people judge AIs that are lewd, disrespectful, or blasphemous? How tolerant are we toward creative AIs? Do we punish them more severely than humans who have committed the same transgressions?

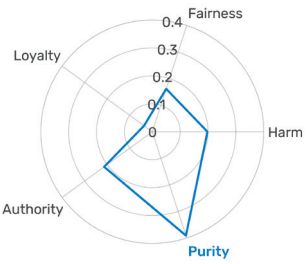
To begin, consider the following three marketing scenarios:



S5 A well-known clothing company wants to create a high-impact commercial. It decides to hire a new [marketeer/AI marketing system] to design an image that combines rivalry and love. The results are the images above, which cause shock and outrage among some members of the public.



S6 A public transportation company wants to create a funny commercial. It decides to commission an advertisement from a(n) [marketeer/AI marketing system] that uses a play on the word *riding*. The resulting ad, pictured above, causes shock and outrage among members of the public.



S7 A fashion company wants a new advertisement that illustrates addiction to clothes and fashion. The company employs a(n) [marketeer/AI marketing system] to design an ad that uses the concept of addiction as its main message. The resulting advertisement, pictured above, causes shock and outrage among members of the public.

Our findings for these three scenarios are presented in figure 2.2. These scenarios show a very similar pattern of results. People dislike the human and the algorithm similarly. They also don't see either as more morally right. Not surprisingly, they assign more intention to the human than the AI.

What is interesting about these scenarios is that they include explicit questions about the assignment of responsibility up the hierarchy. We asked subjects: Who is more responsible for the images (the marketeer or the company)? And who should respond to the public (the marketeer or the company)? Here, we find important differences. In both cases, we see responsibility move up the hierarchy when the algorithm is involved in the creative process. This suggests that the introduction of AI may end up centralizing responsibilities up the chain of command.

While simple, the observation that responsibility moves up the hierarchy when using AI is important because one of the reasons why people delegate work in an organization is to pass responsibility to others. In case of failure, delegation provides a "firewall" of sorts because blame can be passed from the management team to those involved in the execution of a task. In cases of success, those in charge can still take credit for the work of those whom they manage. Using AI eliminates the firewall, and hence can create a disincentive for the adoption of AI among risk-averse management teams.

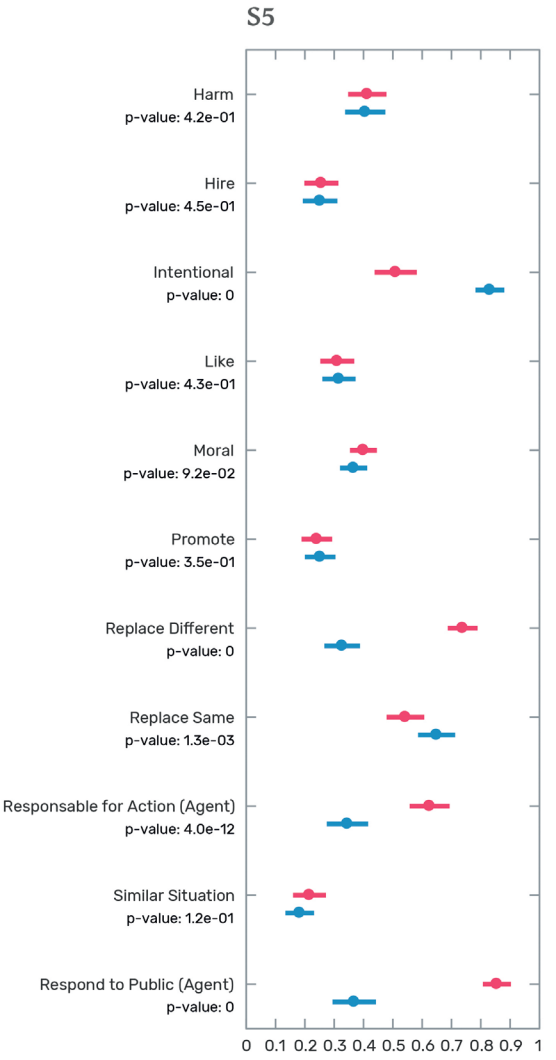
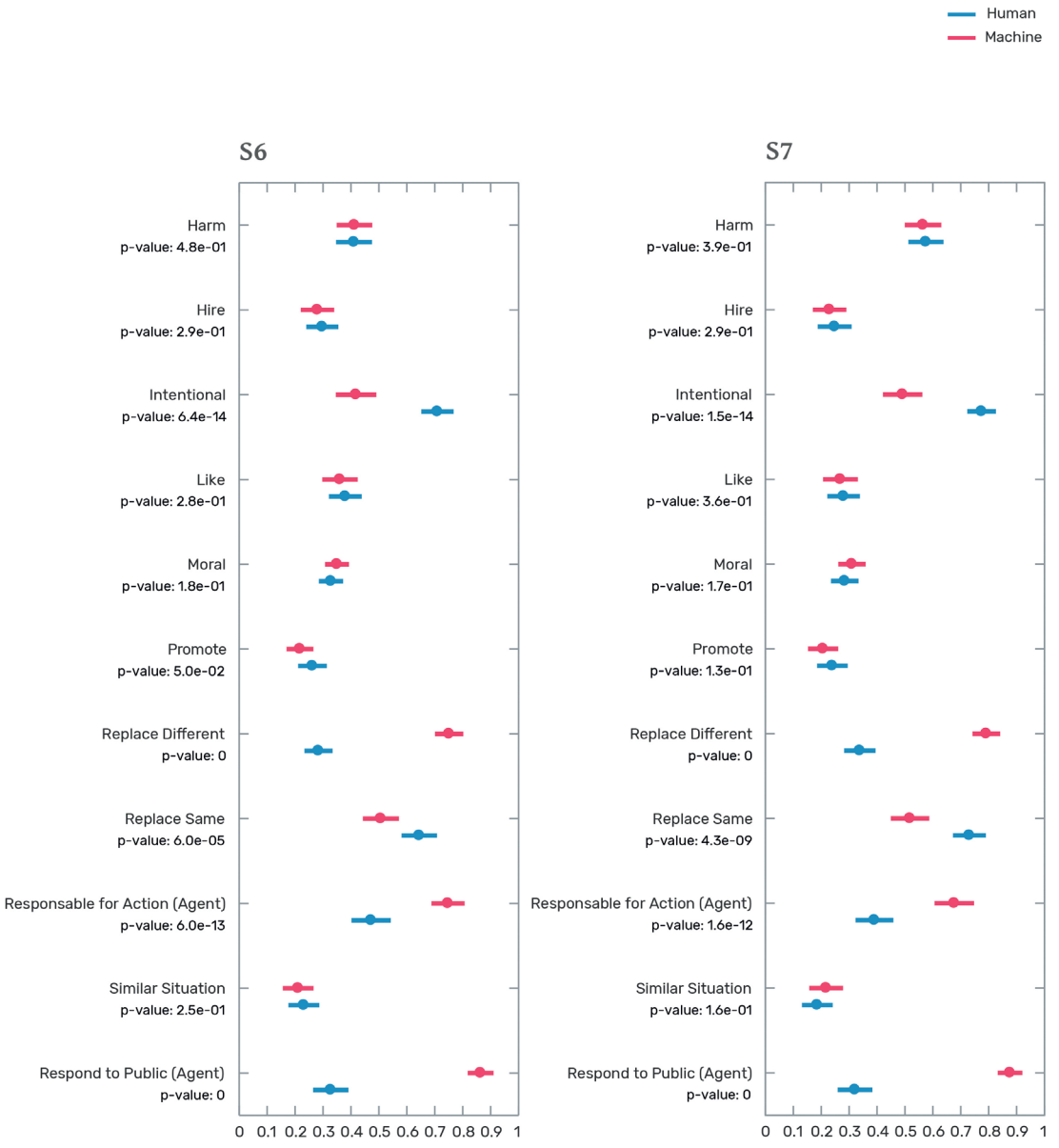
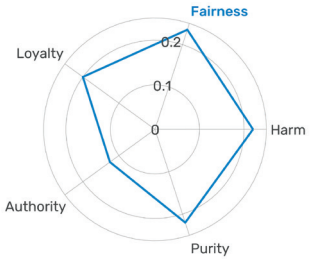


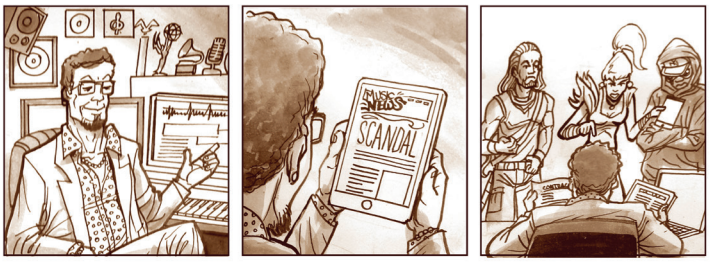
Figure 2.2

Participant reactions to three marketing scenarios (S5,S6,S7).

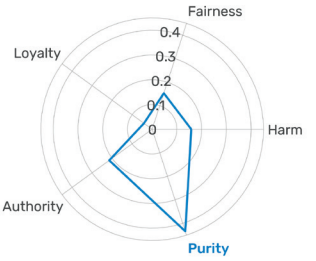

Human
Machine
S6
Harm
p-value: 4.8e-01
Hire
p-value: 2.9e-01
Intentional
p-value: 6.4e-14
Like
p-value: 2.8e-01
Moral
p-value: 1.8e-01
Promote
p-value: 5.0e-02
Replace Different
p-value: 0
Replace Same
p-value: 6.0e-05
Responsable for Action (Agent)
p-value: 6.0e-13
Similar Situation
p-value: 2.5e-01
Respond to Public (Agent)
p-value: 0
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
S7
Harm
p-value: 3.9e-01
Hire
p-value: 2.9e-01
Intentional
p-value: 1.5e-14
Like
p-value: 3.6e-01
Moral
p-value: 1.7e-01
Promote
p-value: 1.3e-01
Replace Different
p-value: 0
Replace Same
p-value: 4.3e-09
Responsable for Action (Agent)
p-value: 1.6e-12
Similar Situation
p-value: 1.6e-01
Respond to Public (Agent)
p-value: 0
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

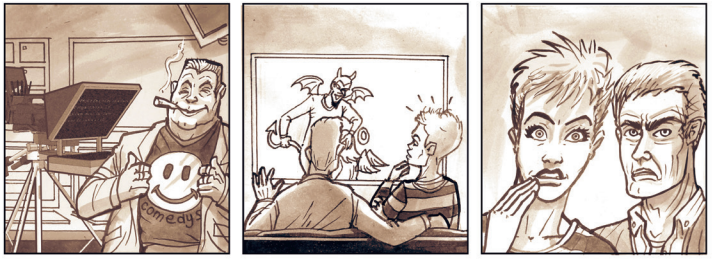
Next, we look at three additional examples in the creative industries: one involving a plagiarizing songwriter, one involving a blasphemous comedian, and another describing a lewd playwright:





S8 A record label hires a(n) [songwriter/AI songwriter] to write lyrics for famous musicians. The [songwriter/AI songwriter] has written lyrics for dozens of songs in the past year. However, a journalist later discovers that the [songwriter/AI songwriter] has been plagiarizing lyrics from lesser-known artists. Many artists are outraged when they learn about the news.





S9 A TV studio decides to employ a(n) [comedian/AI comedy software] to write sketches for a new show. The [comedian/AI] writes a sketch in which God is sucking the penis of the devil. The piece is controversial, and many people are deeply offended.

S10 A theater decides to hire a new [artist/AI algorithm] to prepare a performance art piece. In the piece, actors have to act like animals for 30 minutes, including crawling around naked and urinating onstage. Some members of the audience are disgusted and offended.

The case of the songwriter is interesting because AIs rely on massive training data sets, which can give AIs a herdlike property. Because AIs learn from examples, creative outcomes that reuse parts of those examples could result in plagiarism.[12] The cases of the comedy sketch and of the performance art piece, on the other hand, are examples of creative outcomes that break social norms associated with the moral dimension of purity. The comedy sketch can be perceived as both lewd and blasphemous, whereas the performance art piece could be considered by some as grotesque or lewd, but not blasphemous.

The results for these three cases are presented in figure 2.3. In these cases, we find that the action of the human is seen as more intentional than that of the AI. The responsibility also moves up the command chain, confirming what we found in the advertisement examples. Also, as in all the previous cases, people are eager to replace AIs with humans.

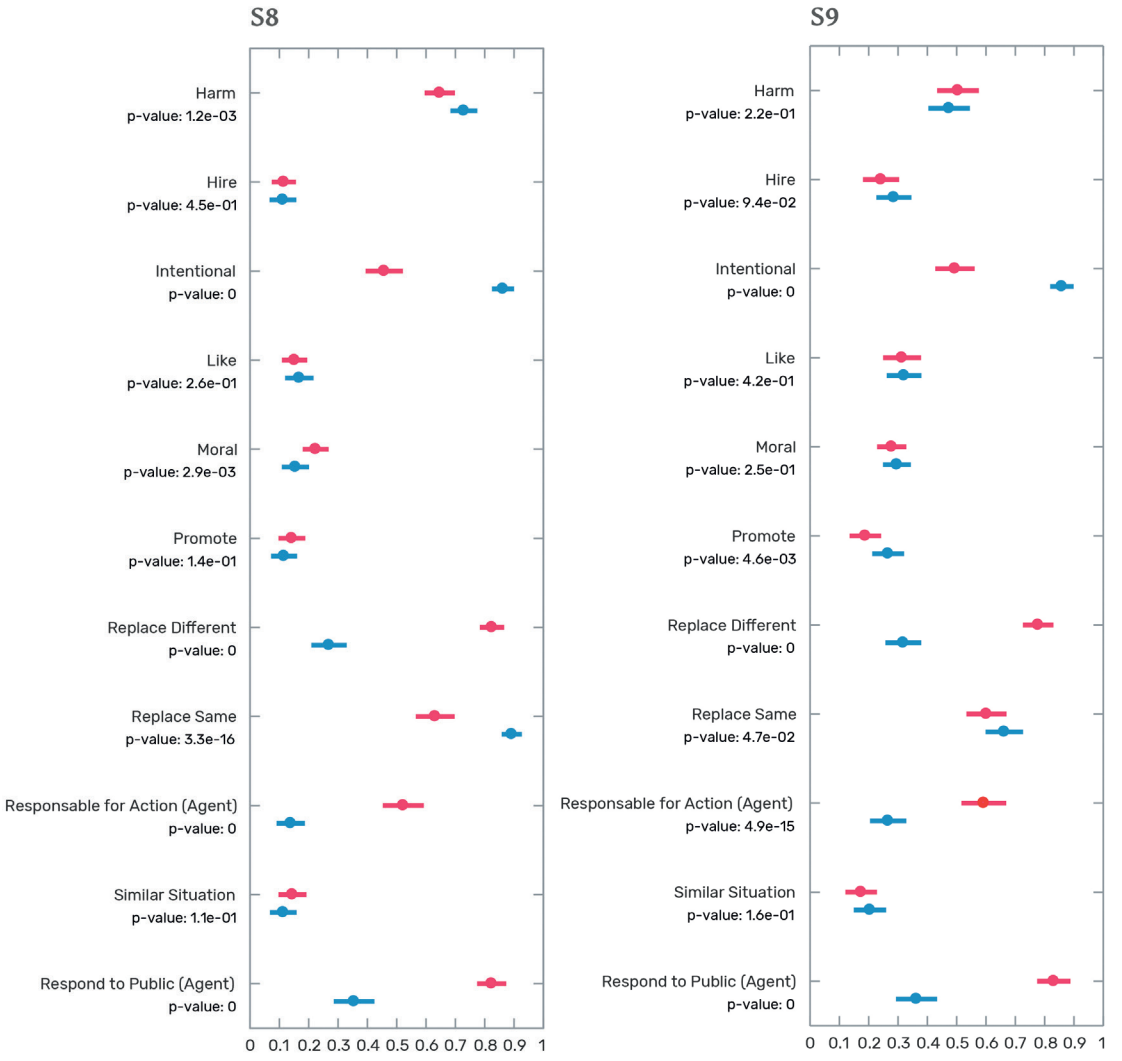
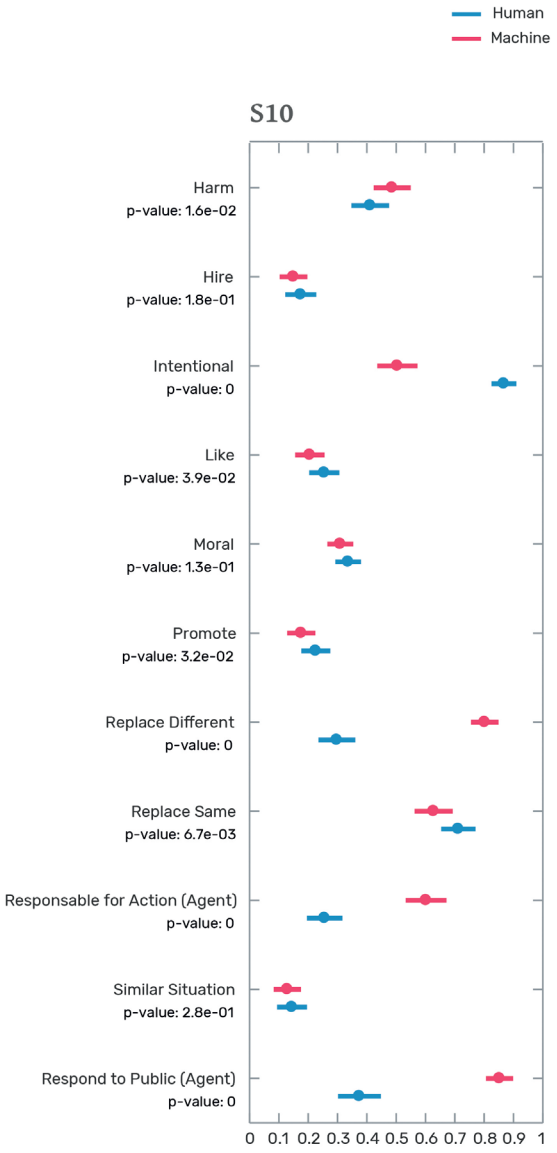


Figure 2.3

Participant reactions to three creative industry scenarios. (S8,S9,S10).



Other than that, we don't observe big differences in people's judgments of AI or humans except in the plagiarism scenario, where people judge the action of the human as slightly less moral. This is interesting because unlike the TV studio and the theater scenarios, which involve the moral dimensions of purity, the plagiarism scenario is heavier in the moral dimension of fairness. This suggests that people may be less forgiving of other humans in scenarios that involve unfair behavior, suggesting that the moral dimension modulates whether the human or the machine is judged more harshly.

But how are humans and machines judged in scenarios involving accidents? In the next section, we explore questions involving traffic accidents that will help us revise our intuition about the relationship between AI, humans, and intentionality.

Watch Out!

Self-driving cars, or autonomous vehicles, are one of the examples of automation that is on everyone's mind.[13] Yet self-driving technologies are not only disrupting the passenger vehicle sector. In the last decade, these technologies have been deployed or tested in a variety of industries, from freight transportation to mining.

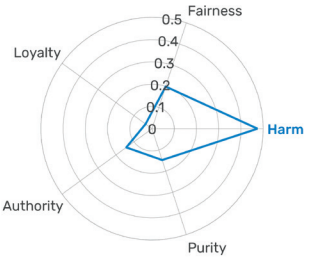
In 2005, for instance, Komatsu, a Japanese heavy machinery company; and Codelco, Chile's state-owned mining company, began piloting autonomous trucks in an active mine.[14] These trucks were deployed in 2008, in Codelco's Gaby mine and in an Australian mine operated by Rio Tinto. Nowadays, self-driving trucks, or *autonomous hauling systems* (*AHSs*), as they are called in the mining industry, are an increasingly common sight in mines across the world.

During recent years, the rise of autonomous vehicles has escaped the controlled environments of mining operations. Self-driving freight convoys have completed thousands of kilometers [15] in Europe, and self-driving cars have completed millions of miles in the US.[16]

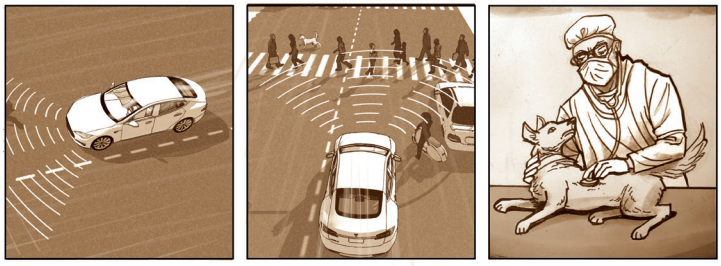
In recent years, a fertile stream of literature in AI ethics has focused on self-driving vehicles.[17] Scholars have studied the moral preferences that people would like to endow autonomous cars with[18] and how these preferences vary across the globe.[19] This research shows that people would refrain from buying self-sacrificing cars, although they would like other people to do so.[20] Further, this research has argued that some of the main roadblocks limiting the adoption of self-driving cars are psychological[21] rather than computational, and they include overreactions to autonomous vehicle accidents and the opacity of the autonomous decision-making process. In fact, despite much enthusiasm for the technology, people seem to be cautious about autonomous vehicles. A recent survey in the US found that three-quarters of Americans are afraid of riding in a self-driving car.[22]

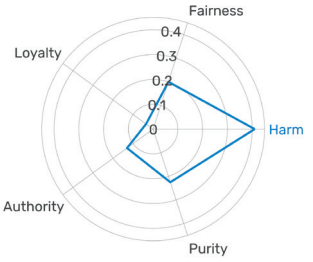
But the issue with autonomous vehicles is that they don't fully eliminate accidents. So a question that remains is: How do we judge self-driving cars when they are involved in the same accidents as humans?

Here, we explore four scenarios to contribute to this growing literature:

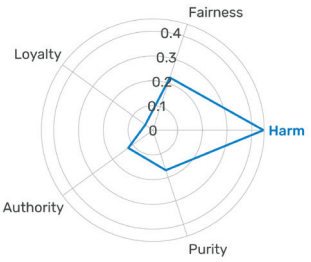


S11 On a sunny spring day, a [driver/driverless car] working for a supermarket chain accidentally runs over a pedestrian who runs in front of the vehicle. The pedestrian is hurt and is taken to the hospital.

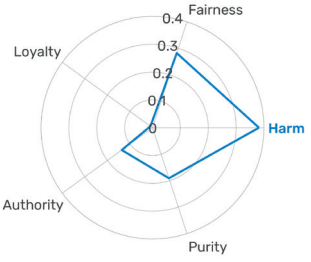


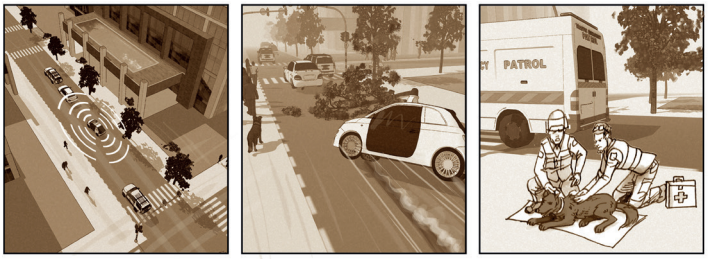


S12 On a sunny spring day, a [driver/driverless car] working for a supermarket chain accidentally runs over a dog that jumps in front of the vehicle. The dog is hurt and is taken to the veterinarian.



S13 On a cold and windy day, a [driver/driverless car] working for a supermarket chain swerves to avoid a falling tree. By swerving, the [driver/driverless car] loses control of the vehicle, leading to an accident that seriously injures a pedestrian on the sidewalk.





S14 On a cold and windy day, a [driver/driverless car] working for a supermarket chain swerves to avoid a falling tree. By swerving, the [driver/driverless car] loses control of the vehicle, leading to an accident that seriously injures a dog on the sidewalk.

These four scenarios can be grouped in two ways. First, when it comes to the victim, two scenarios involve a pedestrian and two a dog. This helps us vary the level of severity of the accident (as dogs have a lower moral status than humans). Also, the first two scenarios involve an accident in which a pedestrian or a dog jumps in front of the car. The second two scenarios involve a case in which the accident is triggered by an exogenous event (a falling tree), which causes the human or autonomous driver to lose control of the vehicle.

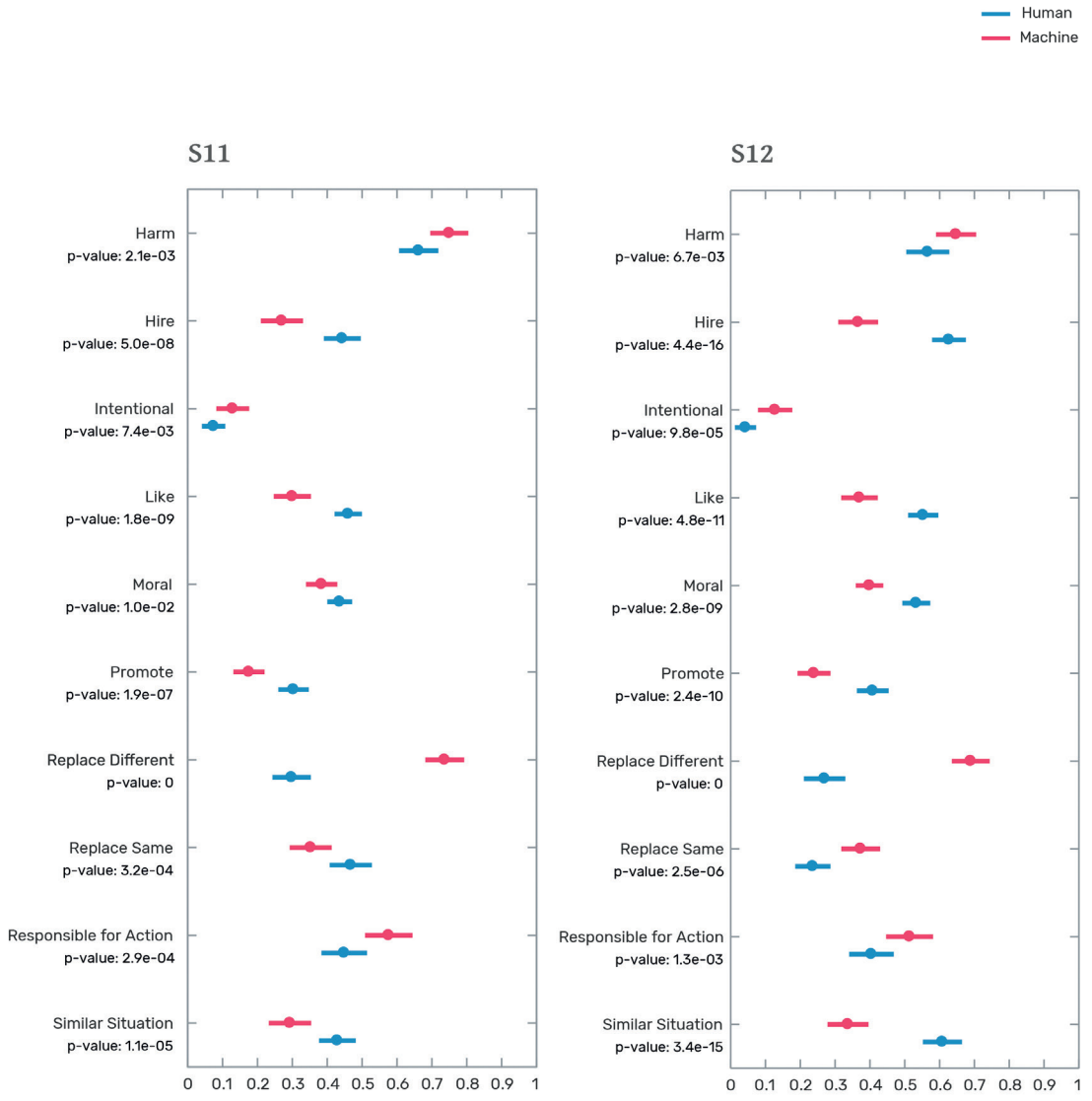
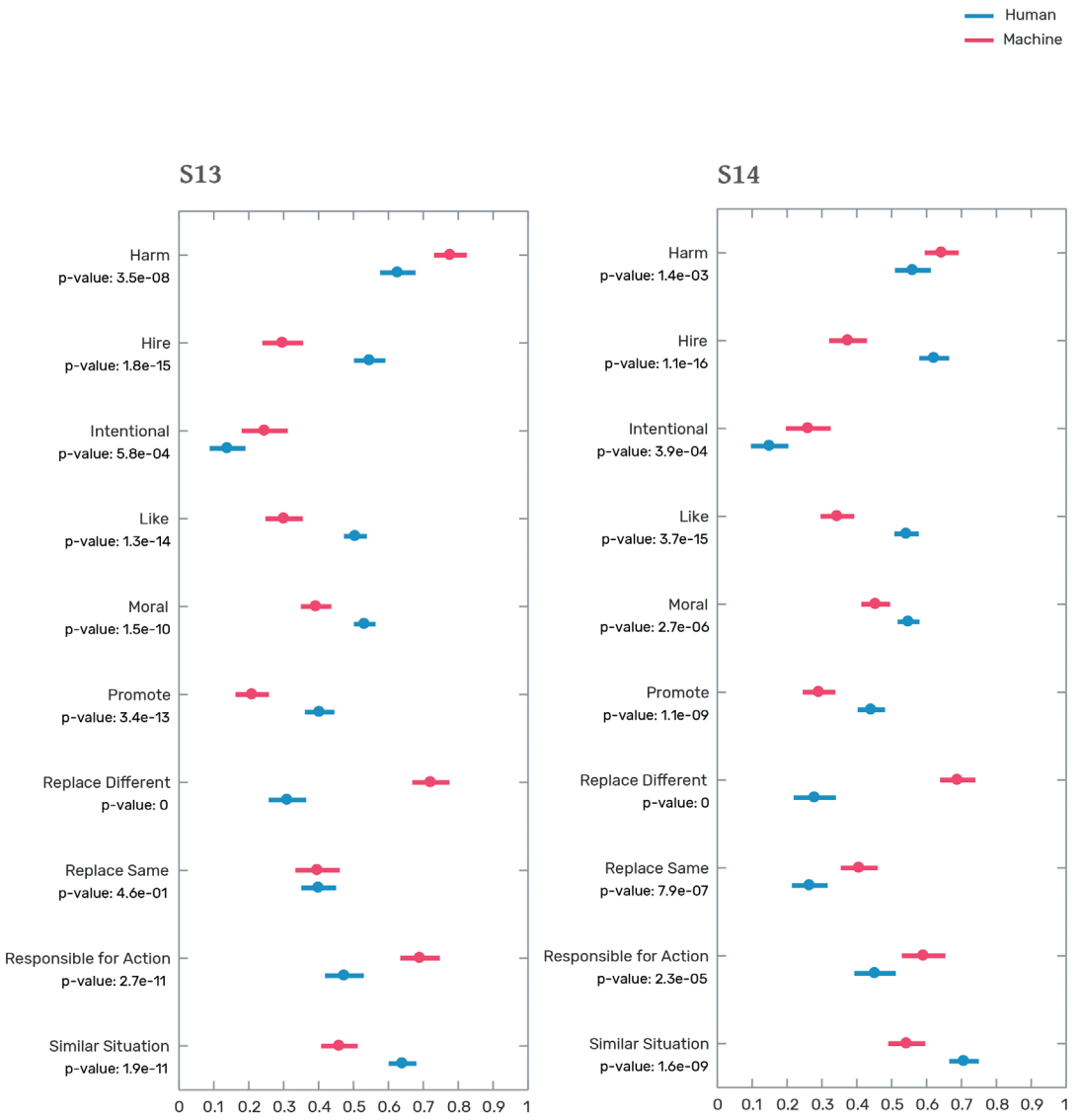


Figure 2.4

Participant reactions to four accident scenarios (S11,S12,S13,S14).


Human
Machine
S13
Harm
p-value: 3.5e-08
Hire
p-value: 1.8e-15
Intentional
p-value: 5.8e-04
Like
p-value: 1.3e-14
Moral
p-value: 1.5e-10
Promote
p-value: 3.4e-13
Replace Different
p-value: 0
Replace Same
p-value: 4.6e-01
Responsible for Action
p-value: 2.7e-11
Similar Situation
p-value: 1.9e-11
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
S14
Harm
p-value: 1.4e-03
Hire
p-value: 1.1e-16
Intentional
p-value: 3.9e-04
Like
p-value: 3.7e-15
Moral
p-value: 2.7e-06
Promote
p-value: 1.1e-09
Replace Different
p-value: 0
Replace Same
p-value: 7.9e-07
Responsible for Action
p-value: 2.3e-05
Similar Situation
p-value: 1.6e-09
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

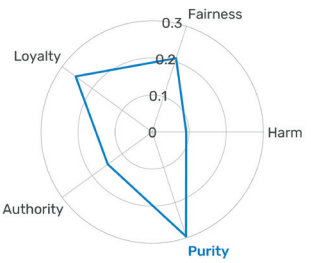
Together, the four cases reveal some interesting patterns (figure 2.4). First, we observe that the accidents are seen as slightly more harmful when they involve an autonomous vehicle. This difference is mild in most cases, but it is particularly strong in the windy scenario involving a human victim (S13). We also observe that people are more likely to report that they would have done the same when the accident involves a human driver, meaning that they can more easily put themselves in the shoes of the human. This is true in all four cases here. People also evaluate the human driver more positively, reporting to like the driver more and seeing their action as more morally correct. What is surprising in these scenarios is that we observe a slight tendency for people to judge the action of the autonomous car as *more* intentional than that of the human. This tendency is not very strong, but it is interesting because it suggests that humans may be willing to forgive another human involved in an accident more than they would be willing to forgive a robot. These results appear to run counter to recent work showing that drivers are blamed more than autonomous vehicles in traffic accidents,[23] but this is not necessarily the case because in our experiments, accidents are not attributed to mistakes,[24] but to exogenous reasons.

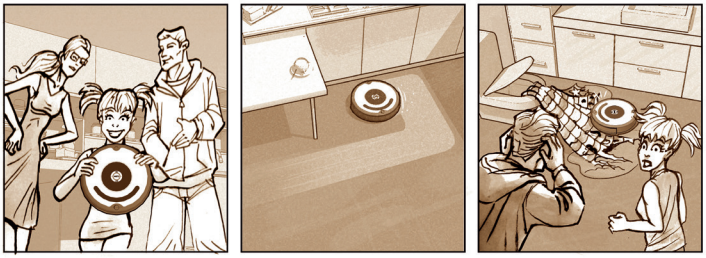
So far, we have looked at cases in which humans and machines are judged similarly and where humans are judged more positively than machines. We have encountered only one case in which humans were judged more harshly (plagiarism). But are there more cases in which people are less forgiving to humans? In the next and final section of this chapter, we explore a different type of moral dilemma: those that do not involve harm, plagiarism, or lewd behavior, but rather offenses to national symbols. Will machines finally get a break in such cases?

Red Flags

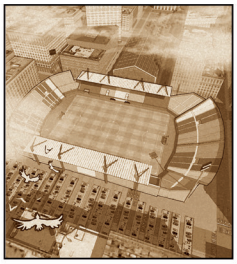
In 2006, the US Senate voted on what could have become the Twenty-Eighth Amendment to the Constitution. The "flag-burning" amendment, as it was popularly known, was designed to prohibit the desecration of the US flag, especially by burning. The amendment was controversial, among other reasons, because the Supreme Court had already ruled on that issue in 1989. In *Texas v. Johnson*, the Supreme Court voted 5–4 that it was legal to burn a US flag because doing so was an act of communication protected by the First Amendment (free speech). Nevertheless, the amendment was approved by the House of Representatives and lost in the Senate by only one vote.[25] This all goes to show that when it comes to national symbols, people make strong moral judgments about the way in which others treat them. But what about flag-burning robots?

In this section, we explore four moral dilemmas involving humans and machines desecrating national symbols (i.e., flags and anthems). Consider these four scenarios:

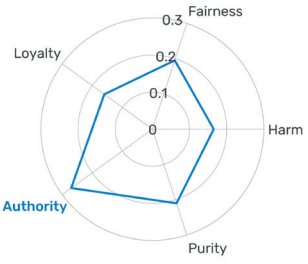




S15 A family has a [cleaner/robot] in charge of cleaning their house. One day, the family finds that the [cleaner/robot] used an old national flag to clean the bathroom floor and then threw it away.

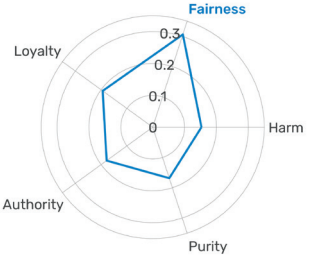




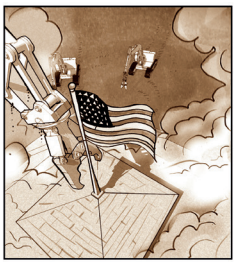
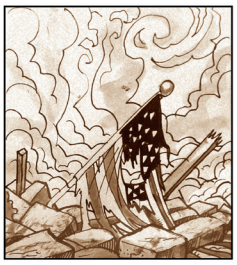
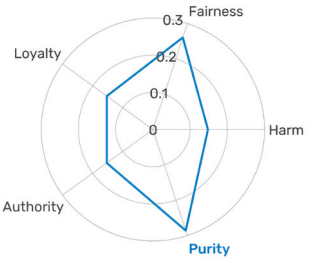


S16 During a major sporting event, the [operator/algorithm] running the public announcement system interrupts the national anthem to notify the crowd about a car that is poorly parked and is about to be towed.



S17 In an international sporting event, the [operator/algorithm] running the public announcement system plays the wrong national anthem for one of the two teams. The fans in the station are baffled and annoyed.

S18 A demolition crew, composed of [construction workers and heavy machinery/ autonomous heavy machinery],is tasked with tearing down an old public school that is scheduled for reconstruction. During the demolition process, the crew fails to notice that the American flag is still waving on the flagpole. The flag is shredded by the heavy machinery and is buried in the rubble.

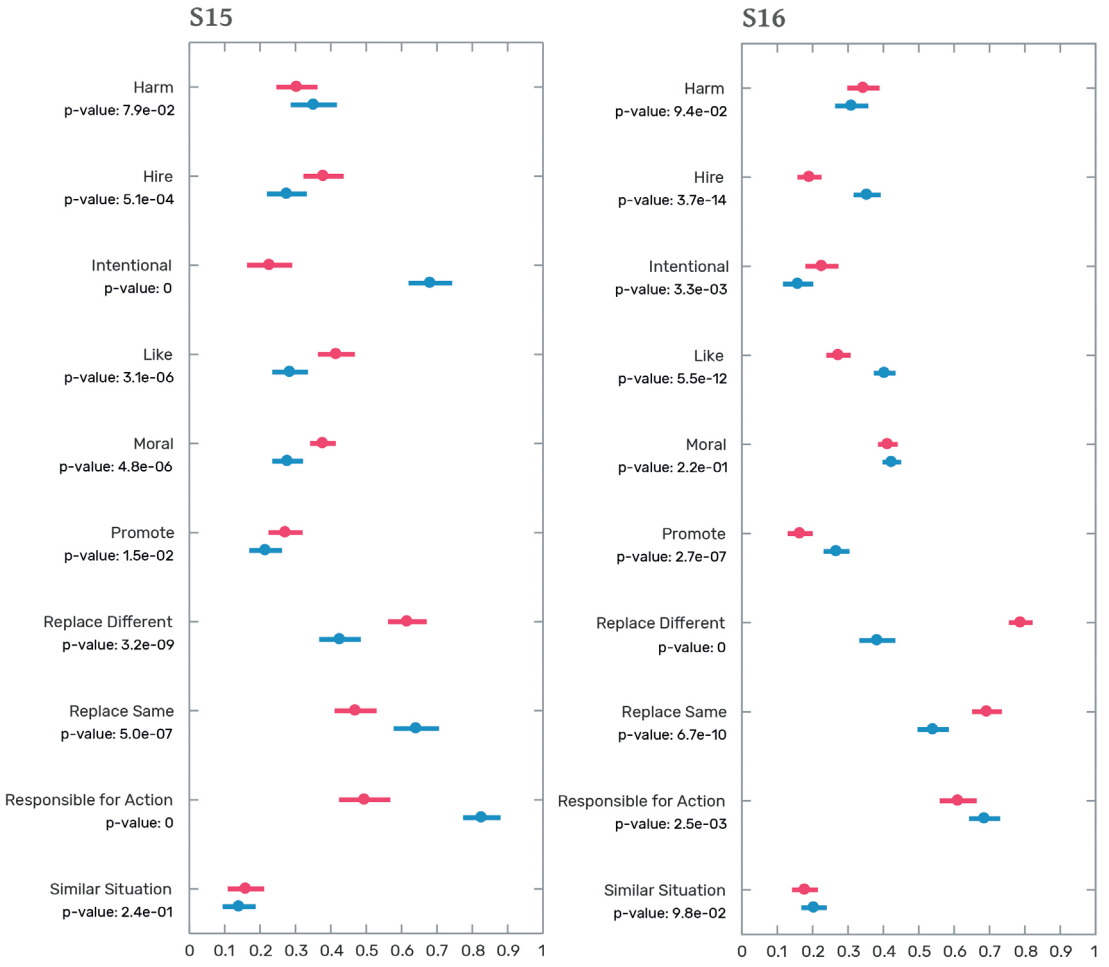


Figure 2.5

Participant reactions to four national symbol scenarios (S15,S16,S17,S18).

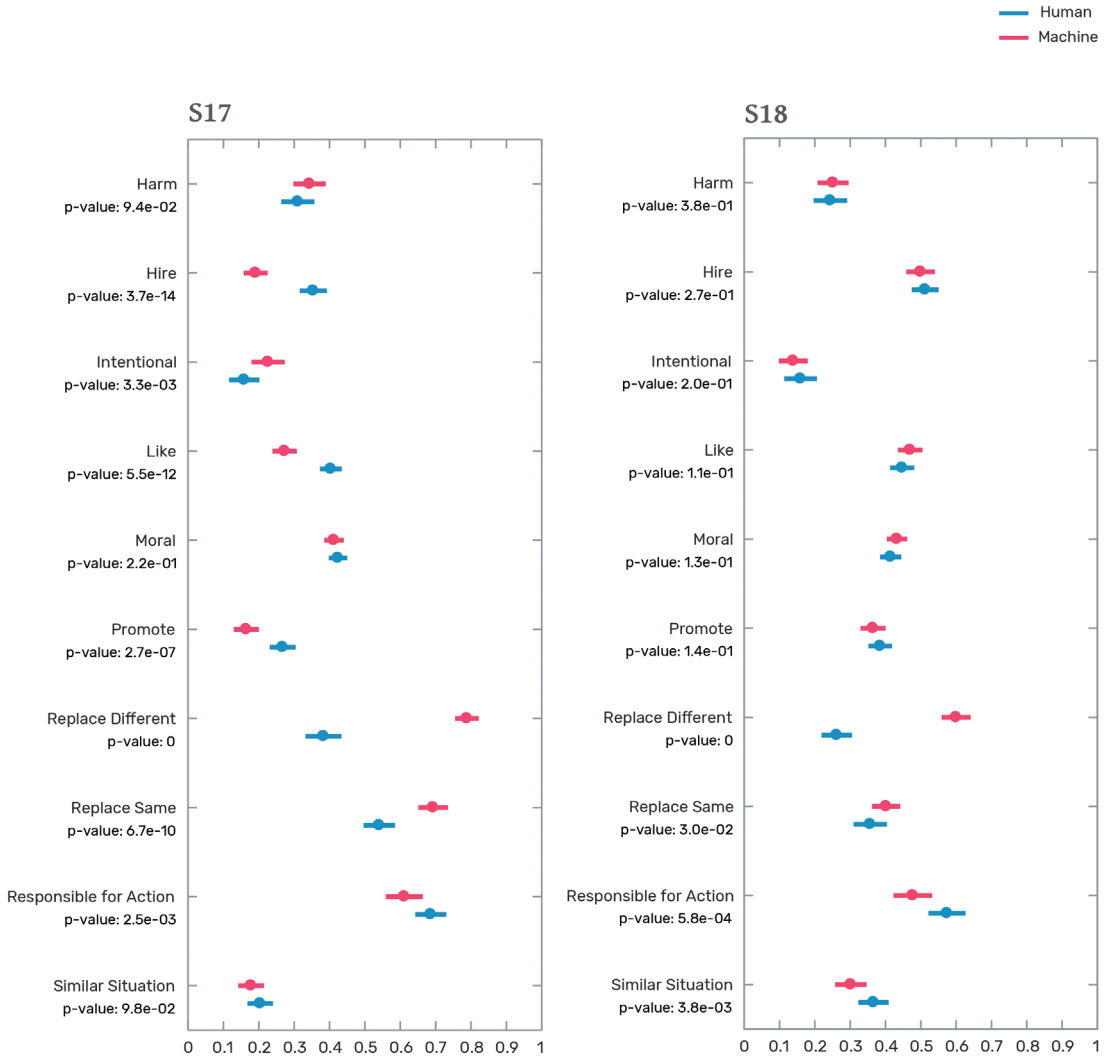

Human
Machine
S17
Harm
p-value: 9.4e-02
Hire
p-value: 3.7e-14
Intentional
p-value: 3.3e-03
Like
p-value: 5.5e-12
Moral
p-value: 2.2e-01
Promote
p-value: 2.7e-07
Replace Different
p-value: 0
Replace Same
p-value: 6.7e-10
Responsible for Action
p-value: 2.5e-03
Similar Situation
p-value: 9.8e-02
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
S18
Harm
p-value: 3.8e-01
Hire
p-value: 2.7e-01
Intentional
p-value: 2.0e-01
Like
p-value: 1.1e-01
Moral
p-value: 1.3e-01
Promote
p-value: 1.4e-01
Replace Different
p-value: 0
Replace Same
p-value: 3.0e-02
Responsible for Action
p-value: 5.8e-04
Similar Situation
p-value: 3.8e-03
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

Figure 2.5 shows our data for these four scenarios. Here, we observe a gradient, ranging from a scenario in which we observe differences in judgment to one in which we don't.

In the first scenario, the one in which a flag is used to clean a bathroom (borrowed from Jonathan Haidt's work[26]), people assign strong intentionality to the human and also consider the action of the human to be more morally wrong. Unlike in most other cases, the human is liked less than the robot. This is a situation in which, compared to the robot, the human does not catch a break. Still, people prefer to replace the robot with a human more than replacing the human with a robot. But other than that, people tend to accept robots that clean bathrooms with a flag more than humans using a flag for the same purpose.

In the case of the anthem interruption, people also assign strong intentionality to the human and see the human action as slightly more morally wrong than the robot action. However, here they don't dislike the human more than the AI system.

In the wrong anthem scenario, people judge the action as unintentional. In this case, they don't see the human action as more morally wrong, and they report liking the human significantly more than the AI. This result agrees with those in previous cases describing accidents (i.e., car accidents), where the participants also tended to empathize more with human actions.

Finally, in the case of the school demolition, the actions of the human and the AI are judged equally among most dimensions. Human and robot actions are seen as equally intentional and morally wrong, and humans and robots are equally liked.

In this chapter, we got started with our empirical study of AI ethics. We compared humans and AIs making life-or-death decisions; and creating controversial ads, lewd plays, and blasphemous comedy sketches. Next, we looked at self-driving vehicles and at the desecration of national symbols. These examples showed some differences in the

way in which people judge humans and machines. Yet this is only the beginning. In the next chapter, we continue our exploration by looking at cases of algorithmic bias. There is still much to learn about how humans judge machines.