A platform for creating a crowdsourced picture of human opinions on how machines should handle moral dilemmas.

BY EDMOND AWAD, SOHAN DSOUZA, JEAN-FRANÇOIS BONNEFON, AZIM SHARIFF, AND IYAD RAHWAN

Crowdsourcing Moral Machines

ROBOTS AND OTHER artificial intelligence (AI) systems are transitioning from performing well-defined tasks in closed environments to becoming significant physical actors in the real world. No longer confined within the walls of factories, robots will permeate the urban environment, moving people and goods around, and performing tasks alongside humans. Perhaps the most striking example of this transition is the imminent rise of automated vehicles (AVs). AVs promise numerous social and economic advantages. They are expected to increase the efficiency of transportation, and free up millions of person-hours of productivity. Even more importantly, they promise to drastically reduce the number of deaths and injuries from traffic accidents.[12,30] Indeed, AVs are arguably

» key insights

- **Machines are assuming new roles in which they will make autonomous decisions that influence our lives. In order to avoid societal pushback that would slow the adoption of beneficial technologies, we must sort out the ethics of these decisions.**

- **Behavioral surveys and experiments can play an important role in identifying citizens' expectations about the ethics of machines, but they raise numerous concerns that we illustrate with the ethics of driverless cars and the Moral Machine experiment.**

- **Data collected shows discrepancies between the preferences of the public, the experts, and citizens of different countries—calling for an interdisciplinary framework for the regulation of moral machines.**

the first human-made artifact to make autonomous decisions with potential life-and-death consequences on a broad scale. This marks a qualitative shift in the consequences of design choices made by engineers.

The decisions of AVs will generate indirect negative consequences, such as consequences affecting the physical integrity of third parties not involved in their adoption—for example, AVs may prioritize the safety of their passengers over that of pedestrians. Such negative consequences can have a large impact on overall well-being and economic growth. While indirect negative consequences are typically curbed by centralized regulations and policies, this strategy will be challenging in the case of intelligent machines.

First, intelligent machines are often black boxes:[24] it can be unclear how exactly they process their input to arrive at a decision, even to those who actually programmed them in the first place.

Second, intelligent machines may be constantly learning and changing their perceptual capabilities or decision processes, outpacing human efforts at defining and regulating their negative externalities. Third, even when an intelligent machine is shown to have made biased decisions,[27] it can be unclear whether the bias is due to its decision process or learned from the human behavior it has been trained on or interacted with.
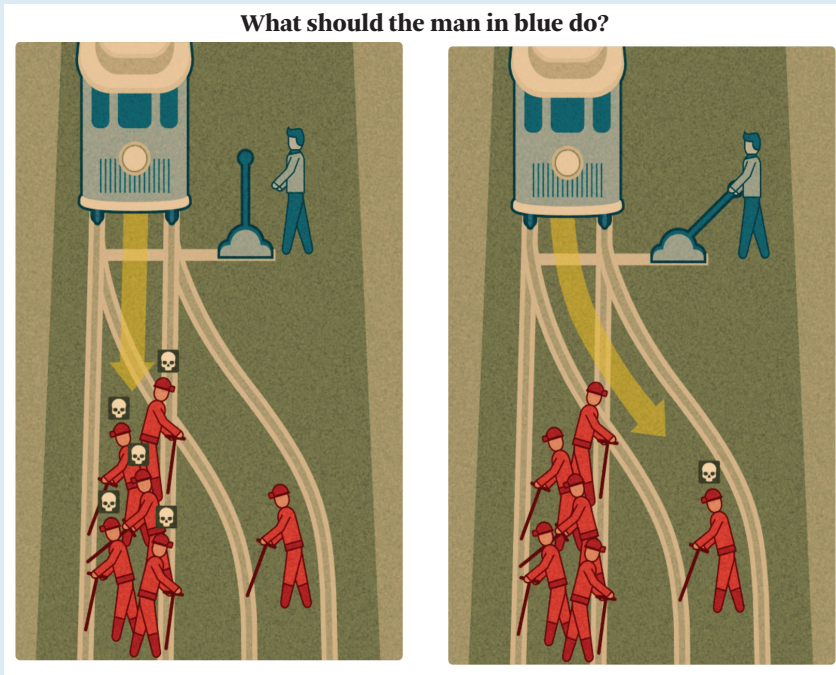
All these factors make it especially challenging to regulate the negative externalities created by intelligent machines, and to turn them into moral machines. And if the ethics of machine behavior are not sorted out soon, it is likely that societal push-back will drastically slow down the adoption of intelligent machines—even when, like in the case of AVs, these machines promise widespread benefits.

Sorting out the ethics of intelligent machines will require a joint effort of engineers, who build the machines, and humanities scholars, who theorize about human values. The problem, though, is that these two communities

Figure 1. A visual depiction of the classic Trolley Problem as displayed in the Moral Machine interface.

A man in blue is standing by the railroad tracks when he notices an empty trolley rolling out of control. It is moving so fast that anyone it hits will die. Ahead on the main track are five people. There is one person standing on a side track that does not rejoin the main track. If the man in blue does nothing, the trolley will hit the five people on the main track, but not the one person on the side track. If the man in blue flips a switch next to him, it will divert the trolley to the side track where it will hit the one person, and not hit the five people on the main track. What should the man in blue do?

What should the man in blue do?

Common criticisms and responses regarding the crowdsourcing of AV ethics using the Trolley Problem method.

Too Naïve	Laypersons' responses to public polls can be biased or ill-informed. Ethical trade-offs must be solved by policy experts, not majority voting.	Policymakers must know about the values most important to the public, so they can either accommodate these values, or anticipate frictions that need be explained.
Too Simple	Real accidents do not involve only two possible actions, and these actions do not have deterministic outcomes.	Highly complex scenarios would only allow for highly specific conclusions. Simplified scenarios zero in on the general principles that guide citizens' ethical intuitions.
Too Improbable	AV-Trolleys are based on very implausible sets of assumptions, and their actual probability of occurrence is too small to deserve attention.	Edge cases can have a massive impact on public opinion, and AV-Trolleys are the discrete form of a very real statistical problem.
Too Early	AV-Trolleys regulations should be avoided at this early technological stage, because their consequences are hard to predict.	Even though it may be too early to regulate about AV-Trolleys, it is the right time to start crowdsourcing citizen preferences.
Too Disconnected	Stated preferences are too disconnected from real actions	The behavior of human drivers is irrelevant to the proposed crowdsourcing task.
Too Distracting	Car makers should focus on making AVs safer, instead of wasting time and resources on crowdsourcing ethical dilemmas.	True, and this is why we need computational social scientist to handle that task.
Too Scary	Overexposing people to AV-Trolleys may scare them away, and be detrimental for their trust in the technology.	This is an empirical question, and our surveys did not find any evidence for such an adverse effect.

are not used to talking to each other. Ethicists, legal scholars, and moral philosophers are well trained in diagnosing moral hazards and identifying violations of laws and norms, but they are typically not trained to frame their recommendations in a programmable way. In parallel, engineers are not always capable of communicating the expected behaviors of their systems in a language that ethicists and legal theorists use and understand. Another example is that while many ethicists may focus more on the normative aspect of moral decisions (that is, what we should do), most companies and their engineers may care more about the actual consumer behavior (what we actually do). These contrasting skills and priorities of the two communities make it difficult to establish a moral code for machines.

We believe that social scientists, and computational social scientists have a pivotal role to play as intermediaries between engineers and humanities scholars, in order to help them articulate the ethical principles and priorities that society wishes to embed into intelligent machines. This enterprise will require elicitation of social expectations and preferences with respect to machine-made decisions in high-stakes domains; to articulate these expectations and preferences in an operationalizable language; and to characterize quantitative methods that can help to communicate the ethical behavior of machines in an understandable way, in order for citizens—or regulatory agencies acting on their behalf—to examine this behavior against their ethical preferences. This process, which we call 'Society in The Loop' (SITL),[25] will have to be iterative, and it may be painfully slow, but it will be necessary for reaching a dynamic consensus on the ethics of intelligent machines as their scope of usage and capabilities expands.

This article aims to provide a compelling case to the computer science (CS) community to pay more attention to the ethics of AVs, an interdisciplinary topic that includes the use of CS tools (crowdsourcing) to approach a societal issue that relates to CS (AVs). In so doing, we discuss the role of psychological experiments in informing the engineering and regulation of AVs,[4,21] and we re-

spond to major objections to both the Trolley Problem and crowdsourcing ethical opinions about that dilemma. We also describe our experience in building a public engagement tool called the Moral Machine, which asks people to make decisions about how an AV should behave in dramatic situations. This tool promoted public discussion about the moral values expected of AVs and allowed us to collect some 40 million decisions that provided a snapshot of current preferences about these values over the entire world.[1]

The Problem with the Trolley Problem

Today, more than ever, computer scientists and engineers find themselves in a position where their work is having major societal consequences.[10,23] As a result, there is increasing pressure on computer scientists to be familiar with the humanities and social sciences in order to realize the potential consequences of their work on various stakeholders, to get training in ethics,[16] and to provide normative statements on how their machines should resolve moral trade-offs. These are new missions for computer scientists, for which they did not always receive relevant training, and this pressure can sometimes result in frustration, instead of leading to the intended ideal outcomes.

The Trolley Problem provides a striking example of the contrast between what computer scientists are trained to do and what they are suddenly expected to do. Scientists working on AVs are constantly asked about their solution to the Trolley Problem, an infamous philosophical dilemma[a] illustrated in Figure 1. At first glance, the Trolley Problem seems completely irrelevant to CS. Its 21st-century version, however, goes like this: An AV with a brake failure is about to run over five pedestrians crossing the street.

---

a   The Trolley Problem, together with all its variants,[11,28,29] is ubiquitous in studies of law and ethics. It was traditionally used to test ethical principles against moral intuitions. More recently, the Trolley Problem has been used extensively in moral psychology and neuroscience to explore not how humans should make ethical decisions, but how they actually do so. This literature delivered deep insights into moral cognition, as well as about the contextual factors that influence moral judgment.[8,9,15]

We believe that social scientists and computational social scientists have a pivotal role to play as intermediaries between engineers and humanities scholars in order to help them articulate the ethical principles and priorities that society wishes to embed into intelligent machines.

The only way out is to swerve to one side, crashing into a barrier and killing its sole passenger. What should the AV do?[13,18] What if there are three passengers in the car? What if two of these passengers are children?

The AV version of the Trolley Problem (AV-Trolley, henceforth) has become so popular that computer scientists, engineers, and roboticists are endlessly asked about it, even when their work has nothing to do with it. It has become the poster child in debates about the ethics of AI, among AV enthusiasts, technologists, moral psychologists, philosophers, and policymakers.[3,18,22] Whether or not this prominence is deserved, the AV-Trolley is everywhere, and it is worth looking in detail at the arguments that have been made for (but mainly against) its relevance for the field of AVs, and for the importance of polling citizens about the solutions they might find acceptable (see the accompany table for a summary).

The citizens are too naïve. First, one may question the usefulness of seeking input from lay citizens when dealing with such complex issues as AV ethics. Certainly, using a simple thought experiment such as the AV-Trolley makes it possible to poll citizens about their preferences. But what are we to do with their responses? Is it not dangerous, or even irresponsible, to seek the opinions of naïve citizens whose responses may be biased or ill-informed? We very much agree that regulations of ethical trade-offs should be left to policy experts, rather than resolved by referendum. But we also believe that policy experts will best serve the public interest when they are well informed about citizens' preferences, regardless of whether they ultimately decide to accommodate these preferences.[2] Sometimes, when policy experts cannot reach a consensus, they may use citizens' preferences as a tie-breaker. Other times, when policy experts find citizens' preferences problematic, and decide not to follow them, they must be prepared for the friction their policies will create and think carefully about how they will justify their choices in the public eye. Whether policy experts decide to take a step toward the preferences of citizens, or to explain why they took a step away, they need to know about the preferences of citizens in the first place.

The scenarios are too simple. Is the AV-Trolley too simplistic to be valuable? Real accidents do not involve only two possible actions, and these actions do not have deterministic outcomes. AVs will have many options beyond staying or swerving, and it is not clear they will be able to precalculate the consequences of all these actions with enough certainty. Many factors that would be relevant for real accidents are simply absent in an AV-Trolley scenario. Note, however, that AV-Trolleys are meant to be abstract and simplified, in order to cleanly capture basic preferences. Using realistic crash scenarios would make it difficult to tease out the effect of multiple contributing factors and make it difficult to draw general conclusions beyond the highly specific set of circumstances that they feature. The AV-Trolley can be used to conduct simplified controlled experiments, in which respondents are randomly assigned to different conditions (accident scenarios), in which the scenarios are simpler than what they would be in the real world, and in which everything is kept constant but for the variables of interest.

The scenarios are too improbable. AV-Trolleys are based on a series of assumptions that are extremely improbable. For example, respondents must accept the very unlikely premises that the AV is driving at an unsafe speed in view of a pedestrian crossing, that its brakes are failing, that there is no other way for it to stop, and that the pedestrians just stay there paralyzed. This combination of unlikely assumptions means the probability of an AV-Trolley actually happening is perhaps too small to deserve so much attention. Or is it? Philosopher Patrick Lin has laid down forceful arguments for the relevance of the AV-Trolley, despite its tiny probability of occurrence.[19] Even if we accept that AV-Trolley scenarios are extremely rare, their consequences may be extremely powerful. The few AV crashes that took place so far received massive coverage in the media, way beyond the coverage of all crashes happening the same year, and way beyond the positive coverage of progress in the performance of AVs. Similarly, a single occurrence of a real AV-Trolley crash could have massive impact on the pub-

AVs will have many options beyond staying or swerving, and it is not clear they will be able to precalculate the consequences of all these actions with enough certainty.

lic trust in AVs. Such a low-probability, high-risk event is known as an edge case, and handling edge cases is important for the design of any product. Finally, even if AV-Trolley crashes are very rare, they can help to think about their statistical extension, the *statistical trolley problem*.[5,14,19] In its discrete version, the AV-Trolley asks about a black-and-white, all-or-none situation where people choose who should live and who should certainly die. The statistical trolley problem ultimately involves the same trade-offs, but ones that occur only when billions of decisions about how minor risks should be allocated are aggregated over millions of miles driven. Imagine an AV driving in a middle lane between a truck and a cyclist. Depending on how much of a berth the AV gives either the truck or the cyclist, its behavior results in a shift of risk between itself, the truck, and the cyclist. This creates the problem of deciding which risk transfers are fair or acceptable. Suppose that conventional cars kill 100 people (80 passengers and 20 cyclists). Program A kills only 20 people (15 passengers and five cyclists), and so does Program B (one passenger, 19 cyclists). What would be the morally preferable program? Should 15 passengers die for five cyclists, or should one passenger die for 19 cyclists? This statistical trolley problem is very real, but much more complex than its discrete version. Data collected with the discrete version of the AV-Trolley do not solve its statistical version but provide a useful starting point for experimental investigations of this statistical version.

Stated preferences are too disconnected from real actions. The idea of "crowdsourcing preferences" assumes that stated preferences provide useful evidence about what respondents would actually do when faced with a physical situation with real life-or-death consequences. But previous work has showed that people's stated preferences and their actual actions diverge in many contexts. In this case, studies that put subjects in simulators and prompt them to react, would provide a better measure of the actual preferences of respondents. While we agree with this assessment, we note that the behavior of human drivers is irrelevant to the proposed crowdsourcing task.

The goal of the crowdsourcing task here is not to capture the actual actions, but to capture what humans would believe (from the comfort of an armchair) to be the best course of action. We can certainly do better with AVs than just imitating the reflexes of a stressed human driver in a split-second crash. Since cars can be programmed and humans cannot, cars can be programmed to do what humans would like to do, rather than what humans would actually decide, on impulse, in a split-second car crash.

It is too early to regulate. Even if AV-Trolley crashes may have major consequences for public trust, they still belong to a rather distant future. They involve highly automated, fully autonomous cars that may not be available for a while, whose behavior on the road is still unknown, and whose technology has not matured. For all these reasons, it may be too early to design regulations for AV-Trolleys. This point relates to the "Collingridge dilemma,"[7] which states that with every new technology, there are two competing concerns. On one hand, regulations are difficult to develop at an early technological stage because their consequences are difficult to predict. On the other hand, if regulations are postponed until the technology is widely used, then the recommendations come too late. In the case of AV-Trolleys, it would seem the ethical debate started well before the technology would be actually available, which means it might be premature to regulate just now. However, it is not too early to inform future regulators about the preferences of citizens. Perhaps right now is not the time to establish rules—but it is the right time to start crowdsourcing preferences, especially when this crowdsourcing effort might take several years.

The debate is too distracting. Car makers are in the business of making safe cars, not in the business of solving age-old ethical dilemmas. By burdening them with the AV-Trolley, the criticism goes, we distract them from their real mission, which is to maximize the safety of AVs, and bring them to the public as soon as possible. This will be better achieved by directing their resources to safety engineering, than to philosophical musing or moral psychology. This is absolutely true, and this is why we be-

lieve that computational scientists have a critical role to play in crowdsourcing machine ethics, and in translating their results in a way that is useful to ethicists, policymakers, and the car industry. The burden must be shared, and computational social scientists are best equipped to handle this crowdsourcing of ethics. Not to mention it is highly implausible the car industry would ever compromise car safety in order to invest in philosophy.

The crowdsourcing is too scary. One main objective of crowdsourcing the ethics of AVs is to find the best possible alignment between regulations and citizen preferences—and a major reason for doing so is to improve trust and social acceptance of AV technology. But crowdsourcing AV ethics using AV-Trolleys could be counterproductive in that respect, since it focuses the attention of the public on scary, improbable edge cases. This is a serious concern, but also an empirical question: Is it true that exposure to AV-Trolleys adversely affects public trust, excitement, or general attitude toward AVs? Our team tested this possibility with both a correlational approach (measuring the link between prior exposure to AV-Trolleys and attitude toward AVs) and a causal approach (measuring the effect of a very first exposure

to AV-Trolleys) and found no statistical evidence for any adverse effect of the exposure to AV-Trolleys.[6] People may not like some specific solutions to AV-Trolleys, but they do not react negatively to the problem itself.
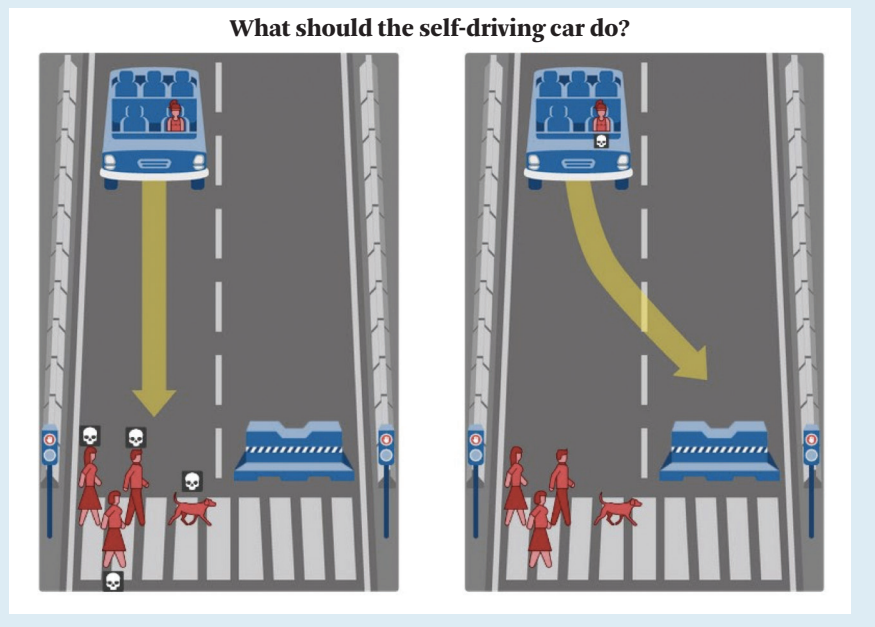
Moral Machine

Having made it clear our support of the use of AV-Trolleys for crowdsourcing the ethics of automated vehicles, the reason for this support, and the limitations of this crowdsourcing exercise, we now describe the platform we created for this purpose, and the data it allowed us to collect. In June 2016, we deployed Moral Machine (MM), a platform for gathering data on human perception of the moral acceptability of decisions made by AVs faced with choosing which humans to harm and which to save. MM fits the specifications of a massive online experimentation tool, given its scalability, accessibility to the online community, and the random assignment of users to conditions. Another purpose to the platform is the facilitation of public feedback, discussion of scenarios and acceptable outcomes, and especially public discussion of the moral questions relevant to self-driving vehicles, which was previously scarce.

The central data-gathering feature is the Judge mode, illustrated in Figure 2.

Figure 2. Moral Machine-Judge interface.

A pictorial representation of a dilemma faced by an AV. If the AV continues ahead it will hit and kill a group of pedestrians, including three adults and a dog, crossing on a red light. If the AV swerves, it will hit a barrier and result in the death of its sole passenger, a female athlete.

What should the self-driving car do?

In this mode, users are presented with a series of 13 moral dilemma scenarios, each with two possible outcomes. The MM restricted scenarios to just two outcomes and did not, for example, offer the solution to drive more slowly and stop safely. This was done on purpose, to ensure participants would have to face difficult ethical decisions, without being able to select a completely satisfying resolution. While this methodological choice was justified in the specific context of the MM project, safe driving and appropriate speed do constitute critically important issues for the broader debate about the ethics of AVs.

The scenarios are generated using randomization under constraints, chosen so that each scenario tests specifically for a response along one of six dimensions (age, gender, fitness, social status, number, and species). Each user is presented with two randomly sampled scenarios of each of the six dimensions, in addition to one completely random scenario (that can have any number of characters on each side, and in any combination of characters). These together make the 13 scenarios per session. The order of the 13 scenarios is also counterbalanced over sessions. In addition to the six dimensions, three other dimensions (interventionism, relation to AV, and legality) are randomly sampled in conjunction with every scenario of the six dimensions. Each of the 13 scenarios features combinations of characters from a list of 20 different characters.

Upon deployment in 2016, the MM website got covered in various media outlets and went viral beyond all expectations. Accordingly, the website's publicity has allowed us to collect the largest dataset on AI ethics ever (40 million decisions by millions of visitors from 233 countries and territories to date).

The results drawn from the data collected through MM were published two years ago.[1] The study reports two main findings: First, among the nine tested attributes, three attributes received considerably higher approval rate than the rest. These are the preference to spare humans over pets, the preference to spare more characters over fewer characters, and the preference to spare the younger humans over the older humans.

Second, while responses from most countries agree on the directions of the preferences, the magnitude of these preferences are considerably different. And countries' aggregate responses broadly cluster into three main clusters: Western (including a majority of English-speaking, Catholic, Orthodox, and Protestant countries), Eastern (including a majority of Islamic, Confucian, and South Asian countries, and Southern (comprising Latin America and former French colonies). The findings also presented predictive factors of country-level differences. One example is the strength of rule of law in a country being correlated with a stronger preference to spare the lawful.

Providing a full discussion about the policy implications of these findings is beyond the scope of this article. However, we note here a summary of the implications. In 2016, Germany became the first country to draft regulations for AVs. The country formed a committee of experts to draft ethical guidelines for automated vehicles.[20] Comparing the preferences we collected via MM to the German commission report, we notice that while there is some overlap between the opinions of the public and the experts (for example, both agree on sacrificing animals in order to spare human life), there are also key points of disagreement (for example, while the public largely approves of sparing children at the cost of the elderly, the ex-

Figure 3. The ranking of countries according to the average preference to spare the lawful (pedestrians crossing at the "walk" signal, instead of the "wait" signal).

All countries show preference for sparing the lawful at the cost of the unlawful.
The top five countries in terms of readiness index[17] are highlighted in red,
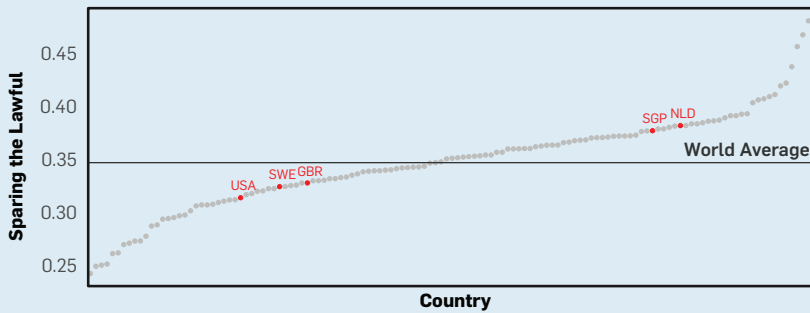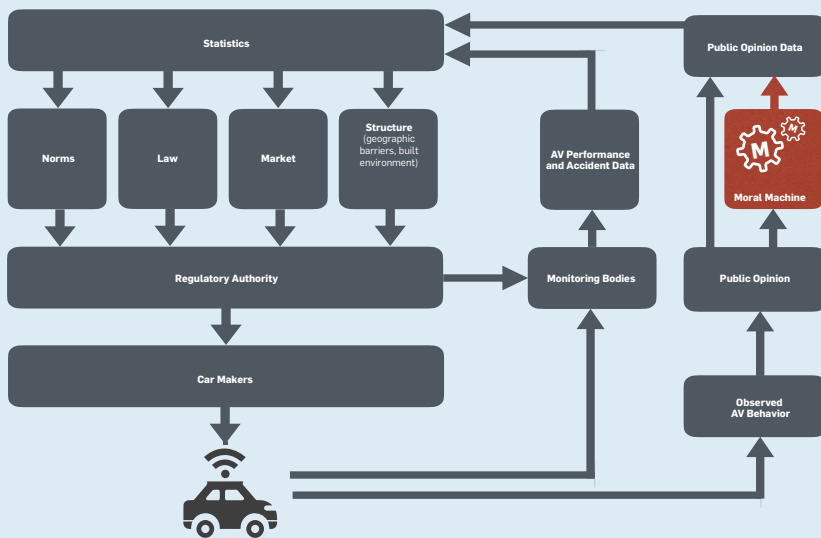and they fall on different sides of the world average for sparing the lawful.

Figure 4. A society-in-the-loop framework for AV regulation.

The model does not represent an actual regulatory system, but it clarifies how a crowdsourcing platform like the Moral Machine fits into the broader regulatory system by providing data on societal norms.

perts prohibit any discrimination based on age). While the experts are not required to cater to the public's preferences when making ethical decisions, they may be interested in knowing the views of the public, especially in cases where the right decision is difficult to discern, and where it may be important to gauge and anticipate public reaction to important decisions.

Clearly, this was the case for Germany. What would be the case for other countries? To date, Germany remains the only country with any guidelines for AVs. Once other countries form their own guidelines, they may end up being similar or different. This leads to our second main finding: Programming ethical decisions in AVs using the same rules is likely to get different levels of push-back in different countries. For example, if AVs are programmed in a way that disadvantages jaywalkers, such AVs may be judged more acceptable in some countries (where the rule of law is stronger) than in others.

The possibility of seeing this happening might manifest itself sooner than we expect. A recent article by KPMG reported on the top countries in terms of readiness for AVs.[17] According to the report, the readiest five countries are the Netherlands, Singapore, the U.S., Sweden, and the U.K. Figure 3 shows that even these top five countries have some disagreement over the magnitude of preference for sparing the lawful. This could mean that a rule such as programming AVs to increase safety for law-abiding citizens at the cost of jaywalkers, while expected to gain high acceptability in the Netherlands and Singapore, may stir anger in the U.S., Sweden, and the U.K.

A Regulatory Framework
As we argued at the beginning of this article, we believe bringing about accountable intelligent machines that embody human ethics requires an interdisciplinary approach. First, engineers build and refine intelligent machines, and tell us how they are capable of operating. Second, scholars from the humanities—philosophers, lawyers, social theorists—propose how machines ought to behave, and identify hidden moral hazards in the system. Third, behavioral scientists, armed with tools for public engagement and

data collection like the MM, provide a quantitative picture of the public's trust in intelligent machines, and of their expectations of how they should behave.[b] Finally, regulators monitor and quantify the performance of machines in the real world, making this data available to engineers and citizens, while using their enforcement tools to adjust the incentives of engineers and corporations building the machines.

We summarize this regulatory architecture in Figure 4, clarifying where crowdsourcing tools can be useful. The Moral Machine project serves as an example of a tool that empowers the public engagement component of our approach to putting 'society in the loop.' It exemplifies interdisciplinary collaboration that combines tools from philosophy, psychology, humanities, computer science and statistics, to inform our quest for a world teeming with increasingly intelligent and autonomous machines that nevertheless behave in line with human values. [C]

---

b We note here that in order to keep the project tractable, the MM experiment had to constrain the possible responses that participants can provide. This is precisely why we do not believe the MM responses are sufficient, on their own, to inform the programming of automated vehicles, which should take into account a variety of perspectives, and the real-world complexity of actual dilemmas of risk distribution. Recent work by Sütfeld et al.[26] suggests the MM results do generalize to different ways of presenting the stimulus, but more work remains to be done on this problem to test the external validity of the findings.

References
1. Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J., and Rahwan, I. The Moral Machine experiment. *Nature 563*, 7729 (2018), 59.
2. Awad, E. and Levine, S. We Should Crowdsource Ethics. In press.
3. Bogost, I. *Enough with the Trolley Problem.* The Atlantic (2018).
4. Bonnefon, J., Shariff, A., and Rahwan, I. The social dilemma of autonomous vehicles. *Science 352*, 6293 (2016), 1573–1576; http://bit.ly/2NyQyUa
5. Bonnefon, J., Shariff, A., and Rahwan, I. The trolley, the bull bar, and why engineers should care about the ethics of autonomous cars. In *Proceedings of IEEE 107*, 3 (2019), 502–504.
6. Bonnefon, J., Shariff, A., and Rahwan, I. The moral psychology of AI and the ethical opt-out problem. *The Ethics of Artificial Intelligence.* S.M. Liao, ed. Oxford University Press, Oxford, U.K., in press.
7. Collingridge, D. The social control of technology. (1982).
8. Cushman, F. and Young, L. Patterns of moral judgment derive from nonmoral psychological representations. *Cognitive Science 35*, 6 (2011), 1052–1075.
9. Edmonds, D. *Would You Kill the Fat Man?: The Trolley Problem and What Your Answer Tells Us About Right and Wrong.* Princeton University Press, 2013.
10. Eubanks, V. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor.* St. Martin's Press, 2018.
11. Foot, P. The problem of abortion and the doctrine of double effect. (1967).
12. Gao, P., Hensley, R., and Zielke, A. A road map to the future for the auto industry. *McKinsey Quarterly*, (Oct. 2014).
13. Goodall, N. Ethical decision making during automated vehicle crashes. *Transportation Research Record: J. Transportation Research Board* 2424 (2014), 58–65.
14. Goodall, N.J. Away from trolley problems and toward risk management. *Applied Artificial Intelligence 30*, 8 (2016), 810–821.
15. Greene, J.D., Sommerville, R.B., Nystrom, L.E., Darley, J.M., and Cohen, J.M. An fMRI investigation of emotional engagement in moral judgment. *Science 293*, 5537 (2001), 2105–2108.
16. Huff, C. and Furchert, A. Toward a pedagogy of ethical practice. *Commun. ACM 57*, 7 (July 2014), 25–27.
17. KPMG International. Autonomous Vehicles Readiness Index: Assessing countries openness and preparedness for autonomous vehicles; https://assets.kpmg/content/dam/kpmg/xx/pdf/2018/01/avri.pdf
18. Lin, P. The ethics of autonomous cars. *The Atlantic* (2013).
19. Lin, P. Robot cars and fake ethical dilemmas. *Forbes* (2017).
20. Luetge, C. The German ethics code for automated and connected driving. *Philosophy & Technology 30*, 4 (2017), 547–558.
21. Malle, B.F., Scheutz, M., Arnold, T., Voiklis, J., and Cusimano, C. Sacrifice one for the good of many? People apply different moral norms to human and robot agents. In *Proceedings of the 10th annual ACM/IEEE Iintern. Conf. Human-Robot Interaction.* ACM, 2015, 117–124.
22. Marshall, A. Lawyers, not ethicists, will solve the robocar 'Trolley Problem.' *WIRED* (May 28, 2017).
23. O'Neil, C. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.* Broadway Books, 2016.
24. Pasquale, F. *The Black Box Society: The Secret Algorithms that Control Money and Information.* Harvard University Press, 2015
25. Rahwan, I. Society-in-the-loop: Programming the algorithmic social contract. *Ethics and Information Technology 20*, 1 (2018), 5–14.
26. Sütfeld, L.R., Ehinger, B.V., König, P., and Pipa, G. How does the method change what we measure? Comparing virtual reality and text-based surveys for the assessment of moral decisions in traffic dilemmas. (2019).
27. Sweeney, L. Discrimination in online ad delivery. *Queue 11*, 3 (2013), 10.
28. Thomson, J.J. Killing, letting die, and the trolley problem. *The Monist 59*, 2 (1976), 204–217.
29. Thomson, J.J. The trolley problem. *The Yale Law J. 94*, 6 (1985), 1395–1415.
30. Van Arem, B., Driel, C., and Visser, R. The impact of cooperative adaptive cruise control on traffic-flow characteristics. *IEEE Trans. Intelligent Transportation Systems 7*, 4 (2006), 429–436

Edmond Awad (e.awad@exeter.ac.uk) is a lecturer in the Department of Economics at the University of Exeter Business School, Exeter, U.K.

Sohan Dsouza (dsouza@mit.edu) is a research assistant at MIT Media Lab, Cambridge, MA, USA.

Jean-François Bonnefon (jfbonnefon@gmail.com) is a research director at the Toulouse School of Economics (TSM-R), CNRS, Université Toulouse Capitole, Toulouse, France.

Azim Shariff (afshariff@gmail.com) is an associate professor at the University of British Columbia, Vancouver, Canada.

Iyad Rahwan (rahwan@mpib-berlin.mpg.de) is a director of the Center for Humans & Machines, Max-Planck Institute for Human Development, Berlin, Germany, and an associate professor at MIT Media Lab, Cambridge, MA, USA.

Watch the authors discuss this work in the exclusive *Communications* video. https://cacm.acm.org/videos/crowdsourcing-moral-machines