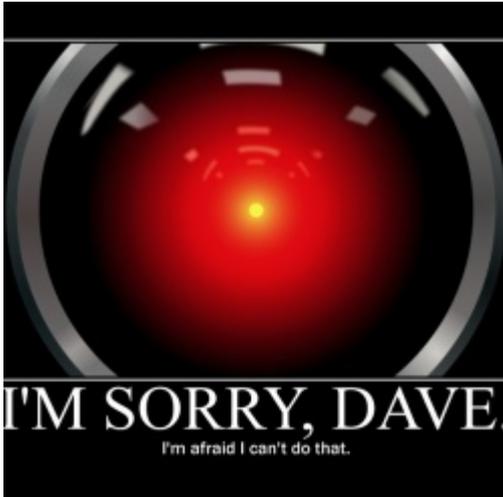# Moral Judgments Are Better Left To Humans Not Machines

By Chuck Dinerstein — September 6, 2018

Courtesy of Cryteria [1]

The world is abuzz with smartphones, smartwatches, smart thermostats – artificial intelligence, by whatever name you choose is upon us. And frequently these systems make demonstrably better judgments than humans, in playing GO, or in optimizing supply chain distributions (a fancy way to describe Amazon's delivery of merchandise). But if the judgment involves a question of morality are the machines better? A recent study looks at our attitudes about computers making a moral decision.

Trolley dilemma

Before jumping to the study, a bit of background may be worthwhile. This frequently cited dilemma comes in many forms. [1] Simply, if you are driving a car and can only avoid a collision with a pedestrian or cyclist by injuring yourself what do you do? Save yourself or a stranger? And how to instruct our cars, to be "selfish" or "altruistic."

In making judgments, humans are guided by two principles, agency and experience. Agency is the ability to think and act, to reason and plan, our expertise; experience in this setting involves the capacity to feel emotions, to empathize. [2] Both are necessary, each is insufficient. Morality is not a purely intellectual domain that machines with great computational ability can master. Morality is shaped by our culture and "deeply grounded in emotions." And computers lack authentic emotions.

The Study

Researchers made use of volunteers, recruited through Amazon's Mechanical Turk, respondings to a series of structured surveys. Each survey enlisting about 240 people and posed scenarios

where a human and a machine made moral choices, in varying fields and with positive and negative outcomes. The participants indicated their preference for "the decider."

- Participants favored humans over machines when given scenarios involving driving, the trolley dilemma, determining parole for prisoners and in making both medical and military decisions.[3]
- The choice of humans over machines was not influenced by knowing the outcomes of their choices and whether the results were positive or negative.
- Human judgments were preferred because of agency and experience were both perceived to be greater with humans than machines.

The researchers then manipulated the surveys to alter agency and experience and reduce the aversion to machine-mediated judgments.

- When expertise was manipulated, an expert machine was chosen over an inexpert human, but when comparable knowledge was present, the human again won out.
- A machine that was more emotionally expressive, more Siri than HAL, did improve the perception of computers as experienced, but the aversion to them making moral judgments persisted.
- When machine advised physician, participants were more willing to have machines involved. But at least a third of participants only wanted physicians to decide.

People use expertise and experience in moral decisions and transfer those considerations to machines; both are necessary, and neither is sufficient. Reducing aversion to moral judgments made by machines is not easy. The perception of expertise must be explicit, over-riding the authority of humans, "and even then, it still lingers." Emotions are critical, core element, to moral-decision making; machines can be made to appear more emotional. People seem to be most comfortable when machines act more as able advisors than primary deciders.

The study has limitations, the domains were restricted, particularly to medicine and drone strikes. It is an online sample and involves moral decisions about third parties, not personal moral choices. And finally, it did not consider responsibility or blame. But it appears, that when making moral judgments we are unwilling to give control to machines when they are equally capable humans to exercise emotional intelligence.

[1] From Wikipedia "You see a runaway trolley moving toward five tied-up (or otherwise incapacitated) people lying on the tracks. You are standing next to a lever that controls a switch. If you pull the lever, the trolley will be redirected onto a side track, and the five people on the main track will be saved. However, there is a single person lying on the side track. You have two options:

1. Do nothing and allow the trolley to kill the five people on the main track.
2. Pull the lever, diverting the trolley onto the side track where it will kill one person.

Which is the most ethical option?"

[2] The preferred term here is moral patiency; is someone or thing deserving of protection. In the

trolley example, it is the driver versus the pedestrian. We make that judgment in many ways, we protect whales and eat fish, some people are vegetarians and others omnivores.

[3] The medical scenario involved paralysis correcting surgery with a 5% risk of death; the military decision was a drone attack with civilian injuries or death.

Source: People are adverse to machines making moral decisions Cognition DOI: 10.1016/j.cognition.2018.08.003