Exploring moral behavior of thinking machines

Abhi Agarwal

Did morality evolve? Frans de Waal centralizes this idea and furthers it by asking how morality evolved. Opposing de Waal on the idea are a small group of biologists. They argue that morality is a construct or an idea that is unique to human beings, and the purpose of morality is to minimize our animalistic instincts. On the other hand, de Waal gives accounts for the idea that morality has evolved continuously and has evolved from Chimpanzees, Bonobos, and Great Apes. He examines different species of animals and points out that primates close to us in the ladder of evolution display a large number of moral traits. Traits such as peacemaking, empathy, sympathy, and many more. For example, if two male chimpanzees are in a quarrel, the female chimpanzees will often mediate and help them settle their quarrel in order to restore the balance in the community.

The smaller group of biologists consists of individuals like Thomas Huxley and George Williams. De Waal, in his book, coined the term Veneer theory to encapsulate their arguments. Veneer theory “assumes that deep down we are not truly moral. It views morality as a cultural overlay, a thin veneer hiding an otherwise selfish and brutish nature" (De Waal, 6). The Veneer theory disregards any connection between morality of humans and the tendencies animals have, and so implies that animals and humans are distinct in this regard. De Waal criticizes this theory in his book primarily for two purposes. The first, the theory lacks proof - de Waal points out that there is no empirical evidence to back the theory. The second, to de Waal it seems unlikely that morality is improved by choice and not through genes. I, personally, side with de Waal. I agree that morality is a product of social evolution - I believe that morality is continuously evolving. Through the proposition of this criticism, de Waal implies that the building blocks of moral agency are apparent within primates (such as Bonobos).

Why is it necessary for some moral behavior to have existed in animals? There is not much of a need of moral behavior when an animal is living alone. However, when animals live within packs or groups then the need for restraint on behavior becomes necessary. One of the truths that is taught to us is that we live in groups because the chance of survival and reproduction are much higher than when living alone. In order for animals to be able to adapt in groups they need to change the primal animalistic nature that they rely on for survival when they live alone. De Wall shows this idea by providing examples of animals that live in groups. In addition, Edward Wilson furthers this idea in his book 'Journey to the Ants'. Wilson describes ant colonies and their successes as a group.

De Waal points out that there are aspects of morality that are unique to human beings. He notes that the ability to weigh, reason, and judge two separate moral decisions and choose an outcome is one of them. The other is the ability to be impartial and spectate a situation. I believe that there is a third. The ability to communicate through an established language is an important aspect that helps us in making our moral decisions. The biggest differentiator between human beings and animals, to me, is the idea that we're able to express things we're thinking about in a formal and clear manner. We're able to have a discourse about the disagreements we have, and debate whether those disagreements are valid or not. These three aspects are critical in the evolution of our morality.

De Waal manages to establish and solidify the idea that there was an evolution of morality, and the roots of morality can be seen in primates. To me, thinking about the evolution of morality is intriguing. We're attempting to build machines that can think and potentially machines that will walk among us. Soon we will progress into developing artificially intelligent machines. This raises a question, can morality evolve to being implemented digitally? How could morality evolve to thinking machines as it did from primates to us?

Firstly, what does it mean for morality to evolve beyond human beings?

What are thinking machines? What exactly are machines that have the ability to make decisions, and mirror our intelligence?

There exists a thought problem in the field of Artificial Intelligence. Hypothetically, I order an AI to build paper clips. The AI, not knowing limitations, would obey its master and would begin creating paper clips. Eventually it would turn the entire universe into paper clips because it doesn't have an understanding of limitations like we do. An AI only knows its task.

What goals would we program a computer to fulfill?

Is morality learned? Does evolution of morality mean we take principles we have learnt of morality and teach/embed that within computers?

Since we could be able to program a computer to make certain deliberations, what should the deliberations be? Should we even make deliberations? There are four possible ways that researchers have thought to be valid options. They are: Direct Specification, Domesticity, Indirect Normativity, and Augmentation. (http://philosophicaldisquisitions.blogspot.com/2014/08/
on-superintelligence-6.html)

References

[1] Bostrom, Nick. *Superintelligence: The Coming Machine Intelligence Revolution.* Oxford: Oxford UP, 2013. Print.

[2] De Waal, F. B. M., Stephen Macedo, Josiah Ober, and Robert Wright. *Primates and Philosophers: How Morality Evolved.* Princeton, NJ: Princeton UP, 2006. Print.