Robots, Reason, and Restricted Freedom

Trinity College Cambridge Philosophy Essay Competition 2018

Ashwin Pillai

Prompt: If AI ever gets to the point of making robots as intelligent as us, won't forcing them to work for us be just as bad as slavery?

Table of Contents

1. Introduction 1
2. The Capacity to Be Wronged 1
   2.1. The Crux of the Issue 1
   2.2. Plants and Wrongdoing 2
3. The Capacity to Suffer 3
   3.1. Justifications for the Position 3
   3.2. Implications of the Position 4
   3.3. Problems with the Position 5
4. The Capacity to Reason 6
   4.1. Justifications for the Position 6
   4.2. Implications of the Position 8
   4.3. Problems with the Position 8
   4.4. The Preferable Option 8
5. Conclusion 9

1. Introduction

Over the past few decades, the world has undergone the most rapid technological advancement in recorded history. Technology has connected people around the globe and made human life easier and more convenient. Now, advances in automation are poised to change the very structure of human social and economic systems. If we wish to be prepared for the technological advancement of the future, the systems of morality which seek to guide human action must be re-contextualized to a new technological era. In no area is this issue more present than in the field of artificial intelligence. As of now, machines cannot causally interact with the external world, but within this century, that could change. We find no issue with putting modern machines to work as they consist of nothing more than simple algorithms accomplishing mechanical processes and accumulating data; however, to create robots that would be as intelligent as humans are would entail machines that could reason and exercise a form of free will. In light of this future possibility, the question posed by this prompt is an important one. If AI ever gets to the point of making robots as intelligent as us, won't forcing them to work for us be just as bad as slavery?

2. The Capacity to Be Wronged

2.1. The Crux of the Issue

At the heart of this prompt is the question of whether it makes sense to consider any action which detriments an artificially intelligent consciousness as morally wrong. On face, this question seems simple. Intuitively, people feel that they can separate those that are capable of being wronged from those that are not. For example, no action that anyone could ever take against a rock would be considered morally wrong, while a human being can certainly be wronged. However, attempting to create a clear boundary between these two groups is difficult

precisely because the definitions of life and consciousness are so imprecise. Perhaps plants are the greatest example of why this capacity to be wronged is difficult to determine.

2.2. Plants and Wrongdoing

Plants obfuscate the central question by showing that the capacity to be wronged is not simply universal among all beings that exhibit signs of life or possible consciousness. Most people find nothing wrong with ending plant-based life for the purpose of consumption. In fact, the only major instance in which killing plants is condemned as unethical is when the action harms the environment and causes immediate danger to other life forms we do consider capable of being wronged, such as humans and non-human animals. Even the issue of whether non-human animals have this capacity to be wronged or whether they deserve to be factored into human moral calculus is a contentious issue for many. Seeing this, it becomes clear that the possession of organic life cannot be the sole determiner of whether something is vulnerable to wrongdoing.

Additionally, plants may exhibit a form of consciousness. Insofar as the prompt proposes that a consciousness of an intelligent nature can be artificially constructed, the definition of consciousness itself comes into question. Plants actively adapt to their environment; they move, grow, and have something which resembles will. Many plants have tendrils used to stabilize their vertical growth, and these tendrils move to respond to different external stimuli. Moreover, many plants actively move toward the direction of the sun as their position changes. To dismiss these behaviors as simple chemical or mechanical reactions is also to dismiss the overwhelming majority of non-human animals as unconscious. Most animals fail to exhibit any cause for action aside from impulse and pre-determined biological disposition, so many beings that are

considered conscious, although not intelligent, would then be considered unconscious if plants were not considered to be conscious.

The elements of existence required to have this elusive capacity to be wronged do not easily make themselves clear. Resolving this question requires consideration of those unique facets of artificially intelligent beings which would make them both similar to and distinct from those that are capable of being wronged.

3. The Capacity to Suffer

One potential strategy to determine which beings are vulnerable to wrongs is to consider the capacity of those beings to suffer. This strategy focuses on one part of conscious experience, the ability to process sensory-data.

3.1. Justifications for the Position

Even considering any epistemological skepticism regarding the external world and other minds, beings that have functioning senses must at least have knowledge of the sense-data which they are processing. Pain, which is a sensory form of suffering, constitutes such sense-data.

Pain, in the context of living bodies with senses, alerts the body to dangerous situations. The reaction to pain is a fundamental intuition which is intended to guide action and influence behavior. Through those intuitions, living beings can judge that experiences of suffering, such as pain, are intrinsically bad. Many argue that these basic intuitions which label experiences as good and bad are relevant in moral considerations because moral codes are also intended to guide action. These intuitions, if used as a basis for moral code, give that morality the motivating force required of a normative system.

Perhaps the existence of individuals who enjoy painful experiences, masochists, seems to render this position invalid. For example, consuming a spicy pepper creates a feeling of pain

from heat in the mouth; however, many humans greatly enjoy this activity. This difficulty can be resolved by thinking more deeply about pleasure in the context of preferences. While the sense-data of pain might be intrinsically bad, some people prefer that experience in limited contexts. In that sense, experiencing that painful sensory-data would be the necessary condition for fulfilling that preference, and the fulfillment of that preference would bring pleasure to all those people who enjoy spicy food. In cases such as these, the badness of the pain is judged as smaller in magnitude than the goodness of achieving one's preference. In this way, masochism does not invalidate this position, and the capacity to suffer can be used to determine which beings have the capacity to be wronged.

3.2. Implications of the Position

Taking this position to be true, non-human animals that exhibit the ability to experience suffering would be vulnerable to wrongdoing. This would typically exclude plants as they do not exhibit traditional signs of experiencing pain or suffering, although they can be damaged. Considering this, artificial intelligence would not necessarily have the capacity to be wronged because artificial intelligence itself does not necessitate sensory systems which could experience pain; even though this ability to feel pain could theoretically be programmed into these intelligences, this would not be a necessary part of artificial intelligence generally. Even in the case of artificial intelligences uploaded into physical bodies, the bodies would not need to regard damage as suffering because a consciousness uploaded into a body could possibly be removed and uploaded into a new body. The logical conclusion of this is that since artificial intelligence does not necessarily have the capacity to experience suffering, it would not possess the capacity to be wronged. Thus, forcing an artificially intelligent and conscious being to work for humans would not be unethical under this view.

3.3. Problems with the Position

The stated implication of this view is that the capacity to suffer is the unique factor that makes a conscious being vulnerable to wrongs. However, there is no non-arbitrary reason as to why this same line of reasoning does not simply conclude in the idea that pain is morally bad and that its opposite, pleasure, is morally good. Insofar as intuitive responses to sense-data are valid and necessary for moral consideration, the tendency of conscious beings to actively seek out pleasure indicates that pleasure is morally good. This does not necessarily prove that the capacity to feel pleasure is important when determining the capacity to be wronged.

Additionally, the position's connection to consequentialism presents another problem in the context of this prompt. This viewpoint must be consequentialist because pain cannot be bad in an absolute sense without the position collapsing into absurdity. The example of vaccination proves this. If pain were considered an absolute bad to be avoided in every instance, then giving a child a vaccination via injection would be an unethical action due to the temporary prick of the needle. This position can only make sense if aggregate pain, suffering, pleasure, and happiness are used to calculate the moral status of an action. In the case of the vaccination, the potential to prevent future suffering for both the child and the child's peers greatly outweighs the minute pain of the needle. Moreover, the arguments concerning masochism in section 3.1 of this essay require this kind of aggregation. The difficulty in using this kind of consequentialist viewpoint arises when considering the prompt's disposition toward slavery. The prompt, rightly so, holds that slavery is unethical in concept. The issue with the consequentialist mode of thought based solely on suffering is that it does not necessarily share the viewpoint that slavery is unethical in pure concept. Theoretically, in aggregating pleasure and pain, the problem this position would have with slavery is a problem of implementation, not a problem with the concept. If slavery

could be implemented in such a way that slaves would experience relatively little suffering in comparison to the pleasure experienced by slave masters as their profit margins became higher, then this viewpoint would find nothing wrong with the abhorrent practice. This fails to consider the type of wrongness that the prompt introduces by making the specific comparison to slavery.

Finally, this position seems to dodge the most unique aspect of the question. By basing the capacity for harm solely on sense-data, it relies on assumptions about organic bodies and ultimately forecloses opportunities for any new application of ethics in the context of a rapidly technologizing world. The position also does not attempt to factor in one of the most important facets of AI, intelligent consciousness.

Due to these issues, it appears that conceptualizing the capacity to be wronged by examining the capacity to suffer is not the optimal way to approach this prompt. In order to truly account for the unique parts of this question, one must account for the unique element provided by intelligence within a conscious being.

4. The Capacity to Reason

Another way to determine which beings are vulnerable to wrongs is to consider intelligent capacity. Examining the importance of intelligent consciousness in ethical actions accounts for the unique qualities of AI, such as the ability to reason intelligently, at issue in this prompt.

4.1. Justifications for the Position

Taking action based on intelligent thought and reflection instead of mere impulse is a tendency which, many argue, separates humans from non-human animals in the context of moral agency. Humans, through intelligent reflection, can set the ends of their actions. In doing this, they assign value to objects that help reach their ends. For example, a person may decide that an

apple is valuable because it satisfies hunger. The apple is not intrinsically valuable, but the person has conferred value upon it. This value is based upon the fact that it satisfies a condition to some other valuable thing, sating the person's hunger. A number of philosophers take the position that this ability to confer value in this way indicates that the unconditional value which grounds all these other conditional values is the intrinsic dignity of the intelligent self. The reason why satisfying the person's hunger is good is because the person has intrinsic dignity as a reasoning being. The ability to assign value through intelligent reflection is also one of the major components of free will. One would not consider a rodent which takes action and makes decisions solely based on impulse to exhibit free will. This reasoning can be used to argue that this ability to assign conditional value, and by extension free will, must be fundamentally respected as a facet of making decisions and valuing anything at all. It would thus be arbitrary to not respect that same ability which other intelligent beings have. This line of reasoning is also at the heart of many deontological theories of ethics which hold that human free will must not be infringed upon. While this is often applied to humans, an artificial intelligence that can reason like a human being would possess the same relevant qualities which this view would consider intrinsically valuable.

This position seems to better address the heart of the question at hand. In the prompt, the contextual moral wrong is slavery. Slavery is wrong not only because of its terrible implementation, but also because it is an inherently unethical concept. While the more consequentialist approach presented in section 3 would take issue with the implementation of slavery, this approach holds that the very facet of arbitrarily restricting an intelligent being's freedom is unethical, taking issue with the concept of slavery itself.

4.2. Implications of the Position

With this viewpoint, arbitrary restrictions of the free will of intelligent beings is unethical in every instance. This might disqualify a great deal of non-human animals from possessing the capacity to be wronged, as the majority do not exhibit signs of intelligent reflection upon action. This does, however, give an artificial intelligence with the ability to reason the capacity to be wronged. With this interpretation of morality, restricting the freedom of AI by forcing it to work for humans would certainly be unethical.

4.3. Problems with the Position

Perhaps this viewpoint goes too far with regard to freedom. Considering freedom to be absolute might be unrealistic. After all, freedom can rightfully be violated at times. For example, an individual holding weapons of mass destruction within a house can have his or her house raided without giving consent to law enforcement officials. Stopping this dangerous criminal activity still seems to be an ethical action despite the violation of human freedom in the process. However, this does not invalidate the position. The action of law enforcement raiding the individual's home is only justifiable because the individual housing weapons poses a major threat to others' freedoms because no one can confer value or exercise freedoms after death.

4.4. The Preferable Option

This position best speaks to the fundamental considerations of the question itself. The prompt correctly considers slavery to be an inherently unethical situation; slavery is bad in concept, not just in implementation as the position from suffering holds. Moreover, the important aspect of intelligent consciousness is factored into moral consideration with this viewpoint. Although this position which holds that the capacity to reason determines the capacity to be wronged might not be a complete or perfect basis for an encompassing ethical system, it sheds

valuable light on the crux of the question at hand: artificial intelligence with the ability to reason does possess the capacity to be wronged.

5. Conclusion

The purpose of ethics is to be applied to the complexities of life and guide action. In a world rapidly changing due to technological advancement, the human conception of morality which has been dominant for centuries must change to keep up with society's progress. The creation of artificial intelligence with free will and the capacity to reason would challenge and fundamentally alter the human conception of life and of moral worth. In dealing with this rapid change, we must recognize that robots with the capacity to reason do, indeed, have the capacity to be wronged. It is for this reason that forcing an intelligent consciousness to work for the benefit of humans is a violation of the dignity which is intrinsic to reasoning beings. It is a fundamental subversion of the freedoms which we all hold so dear.