Should robots have rights? Why or why not?

At its most basic a right can be defined as a moral or legal entitlement to have or do something, used most commonly in the context of legal and ethical protections or responsibilities. For example, the 'right to life' gives us legal and moral status that protects us from being killed and equally gives us an obligation not to kill. Throughout history, we see a progression of the increase of rights shown by documents such as the Universal Declaration of Human Rights and in more recent times the efforts to extend many of these rights to animals, because of this it's inevitable that as robots and artificial intelligence become more advanced and 'human-like' the legal and moral status of robots will enter the foreground of philosophical debate. In this essay, I will argue in favour of a utilitarian approach to how we award rights and how using this approach alongside utilising parallels that can be found in the relationship between the rights of humans and animals it is also justified to give robots rights.

Utilitarianism is the idea that concepts of morality can be reduced down into pleasure and pain which in turn means all moral actions are ones that prevent pain and promote pleasure, using the lens of utilitarianism all laws and the rights associated with them become reducible into pain and pleasure. For example, it's easy to see how the human right to be free from torture is easily reducible, but even more abstract concepts such as the right to freedom of expression can be reduced down to the emotional pain we feel if the psychological need to expression we carry as social animals are denied to us.[1] It is this theory of rights that is most consistent with how we tend to use the notion of rights in reality, for example, treatment of children and the less mentally able are clear indicators of how our notion of rights is not derived from either ability to reason or communicate; both

[1] Human Rights Act (1998)

these groups are less able to rationalise or communicate and yet despite this they still receive the same rights as those more able to communicate and rationalise. In fact, in our recognition that these groups are more vulnerable and therefore more likely to experience pain we award them extra legal protections to prevent this. Regarding to who is deserving of rights Bentham writes that "the question is not, Can they reason? nor, Can they talk? but, Can they suffer?"[2] Essentially Bentham is arguing that this moral principle should be extended to all who poses the ability to feel pain and therefore would need to be considered in utilitarian calculations of pain and pleasure.

Peter Singer later expands upon this idea and how it relates to the moral and legal rights of animals through his popularisation of the idea of speciesism[3] (a term first used by Richard Ryder) arguing that generally the rights and interests of humans will be given more significance compared to the rights of animals despite animals being just as able to feel pain, thus meaning that we are discriminating against animals based on their species similar to how a bigot may discriminate based on ethnicity or skin colour. And so, for a utilitarian whose only concern would be if it's possible for them to experience pain and pleasure just like how ethnicity would not be a relevant factor nor would species. So, therefore there is no logical justification for considering our rights more important than theirs.

Using a similar form of logic if robots have the ability to feel pain then they are deserving of rights, and a denial of these rights to be similarly unjust as speciesism. This possibility of robots being able to feel pain can be justified by the idea of multiple realizability, popularised by Hilary Putnam, through its principle that the

[2] Of the Limits of the Penal Branch of Jurisprudence, Jeremy Bentham (1782)
[3] Gruen, Lori, "The Moral Status of Animals", *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2017/entries/moral-animal/>

same mental state (pain) can be realised by different physical states which mean that in order to feel pain a robot would not be required to have the brain of a human. The most significant argument in favour of multiple realizability is 'the likelihood argument' given by Putnam and best explained by Shapiro[4] if we consider two possibilities:

> (TI): Psychological states are (type) identical to physical-chemical states of the brain.
>
> (MR): Psychological states are multiply realizable.

Then it becomes clear that the second option (multiple realisability) is preferable to the first (Type-Identity theory, which can be considered the other most prevalent theory of mind) because the first option implies that in order to share the mental state of pain all organisms must share the same physical-chemical states of the brain, which science suggests is false. So, from what we know about life on earth we can conclude that the same principle would apply to other kinds of minds, such as in robots, meaning it is probable that multiple realizability is true which allows for the possibility of robots feeling pain.

One of the strongest arguments against multiple realisability (and therefore the strength of this essay's argument) is directed at the empirical evidence given by Putnam using an example from animals as justification for the 'likelihood argument'. Bechtel and Mundale[5] argue that even if the examples given are true this doesn't give any validity to multiple realizability because, they argue, "it is the very similarity (or more precisely, homology) of brain structures which permits us to generalize across certain species" which they argue means multiple realisability fails. As Sungsu

[4] Shapiro, L. A. (2000). *Multiple Realizations. The Journal of Philosophy, 97(12), 635.* doi:10.2307/2678460
[5] Bechtel, W., & Mundale, J. (1999). *Multiple Realizability Revisited: Linking Cognitive and Neural States. Philosophy of Science, 66(2), 175–207.* doi:10.1086/392683

Kim explains in his commentary of their paper[6] homologies are traits that species share due to them being descendants of a common ancestor giving the example of human limbs and bird wings coming from the forelimb of a common ancestor. However, homoplasious traits also exist which are similar traits that have evolved independently from one another for example bird and bats evolved wings independently from one another. Essentially Kim points out that Bechtel and Mundale's argument is irrelevant because it is the homoplasious traits of brains that will provide evidence against multiple realisability just as the homologous traits of human limbs and bird wings cannot disprove the homoplasious traits of bird and bat wings the homologous traits of some brains (e.g., human and primate brains) cannot disprove the possibility of homoplasious brains. If we were able to be provided with evidence of brains that evolved homoplacsiously and ended with similar structures this would imply that multiple realizability was not possible because if it was possible for brains to exist in vastly different forms, then it is probable the homoplasious brains would do so.

Another criticism of more extreme forms of multiple realizability is given by Kim Jaegwon[7][8] pointing out that the successes found in neuroscience would suggest commonalities between the mental processes of other animals and humans, otherwise there would be no point in studying the neuroscience of animals in order to gain a better understanding of human neuroscience. Kim argues that neuroscience is only successful because we can assume some commonalities between different species and that therefore these techniques would be useless

[6] Kim, S. (2002). *Testing Multiple Realizability: A Discussion of Bechtel and Mundale. Philosophy of Science, 69(4), 606–610.* doi:10.1086/344623
[7] Bickle, John, "Multiple Realizability", *The Stanford Encyclopedia of Philosophy (Spring 2019 Edition)*, Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2019/entries/multiple-realizability/>
[8] Kim, J. (1992). *Multiple Realization and the Metaphysics of Reduction. Philosophy and Phenomenological Research, 52(1), 1.* doi:10.2307/2107741

when used with extreme forms of multiple realisability, for example, the 'China Brain' outlined in Ned Block's thought experiment. It can be argued that this argument is invalid in regard to robots being able to feel pain for two reasons. The first is that it assumes that just because neuroscientific methods of research would be pointless on China Brain this automatically means that it is not valid as a mind. A better analysis could be that neuroscience is only applicable to biological minds and therefore extreme forms of multiple realization cannot apply to biological minds, however, the concept of minds as a whole and therefore the ability to experience pain is still valid through multiple realizability. Another argument that could be proposed is that potentially neuroscientific methods of experimentation could apply to the 'brain' of a robot. The technique of positron emission tomography is used to detect changes in blood flow through the emission of radioactive substances if the brain of a robot were structured and programmed in a comparable way it is possible that the change in blood flow caused by pain would be comparable to the measuring of the temperate of certain regions of the robot's 'brain' that had been programmed to be responsible for pain. Thus, meaning that if created in likeness to the brains of humans and other organisms the success of neuroscience and its techniques wouldn't be valid as a criticism. Although this would essentially create an example of trivial multiple realizability as oppose to substantive multiple realizability (a distinction made by Shapiro[9]) in this case this isn't an issue as for this essay as it only seeks to argue that a robot would be able to feel pain, not the validity and significance of multiple realizability as a whole so resorting to the trivial form of multiple realizability isn't an issue as long as it is able to support the argument.

[9] Shapiro, L. A. (2000). *Multiple Realizations. The Journal of Philosophy, 97(12), 635.* doi:10.2307/2678460

In conclusion, because we are morally required to extend rights to everything capable of feeling pain and pleasure as justified by utilitarianism, empirical evidence of how we treat vulnerable groups and the bigotry of speciesism, paired with multiple realizability making it at least feasible that a robot could feel pain this would mean that if such advances in technology were to be made then we would be morally required to give a robot rights.