

Moral status, suffering, and compassion: Towards reconciling human moral priority with animal welfare**Adam Joseph Braus¹****Abstract**

This paper critically examines the principle of equal consideration of interests (ECOI) in the discourse of animal ethics. I claim that within a consequentialist moral order, an entity's moral status comes from working backward from consequences, not solely on intrinsic capacities. Hence, I propose that to judge moral status, one must take into account not only (a) an entity's intrinsic capacity for sentience but also (b) how reliably that entity will reduce suffering overall. Given these two criteria, I explore whether human beings deserve a higher moral status than non-human animals. While there is evidence that *homo sapiens* have evolved a unique instinctual urge to reduce suffering called 'compassion,' humans are not always compassionate. For this reason, I conclude that as a species, human beings do deserve a much higher moral status than other animals; however, the moral status of any individual human being falls upon a much wider spectrum above and below other animals. This framework has the advantage of justifying common moral intuitions about human moral superiority while still obligating humans to exercise compassion. I conclude by discussing the implications of this on the moral status of non-paradigm humans, and its possible import for AI ethics and alignment.

Keywords: speciesism, moral status, Animal Ethics, suffering, compassion, consequentialism

Introduction

The question of moral status—determining which beings deserve our ethical consideration and to what degree—stands at the heart of contemporary discussions in animal ethics.² Traditional approaches have often created a dichotomy: either we accept human exceptionalism, which grants humans categorical moral priority regardless of an animal's capacities, or we embrace a strict human-animal egalitarianism based solely on sentience.³ Here, I propose a more nuanced consequentialist theory of moral status that incorporates both (a) the capacity to suffer and (b) the reliability of reducing suffering as twin pillars for determining moral status.⁴ This approach offers a new and more nuanced way to reconcile human moral priority and moral agency with the obligation to treat all animals and non-paradigm humans with compassion. Adding this additional dimension to moral status stands in contrast to other approaches that are criticized for being able to justify treating non-human animals or non-paradigm, vulnerable, or less socially useful humans without compassion.

This paper has the following structure. In Section 2, I critically examine the current discourse on moral status, especially the principle of equal consideration of interests (ECOI). In Section 3, I propose and defend judging moral status by an entity's capacity to suffer and how reliable they are at reducing suffering. In Section 4, I approach the key question of whether humans deserve moral priority over animals or not. I conclude that humans show such variability in their reliability at reducing suffering that they deserve a variable moral status. Those who

¹ Dominican University of California, San Rafael, CA (USA); email: adam.braus@dominican.edu; ORCID: 0009-0002-4093-2681

² I use the term 'moral status' as a close synonym with 'moral standing' or 'moral consideration'. By this term, I mean (a) a being can be benefited or harmed in morally relevant ways, (b) how one treats this being matters from an ethical perspective, (c) one ought include the being's interests or welfare in one's moral reasoning, and (d) one ought to accept some level of moral obligation or responsibility toward them (Shepherd, 2018).

³ Following Bentham and Singer, I consider the capacity for suffering, having interests, or sentience as largely equivalent in this discourse and refer equally to the capacity to suffer and the capacity for sentience.

⁴ By 'suffering' I mean aversive and unpleasant feelings associated with disruptions to welfare which includes the frustration of preferences, disruption of agency, disruption or damage to personhood, unwanted physical pain, and the like (Singer, 2023; Kauppinen, 2019; McClelland, 2010; Cassel, 1998).

reduce suffering overall deserve a higher moral status, and humans who create greater suffering deserve a lower (even sometimes a negative) moral status. In Section 5, in light of these considerations, I discuss the moral status of hypothetical artificial super-intelligences (ASI). I conclude that, since ASI has behavior not bound by unalterable genetics but by changeable programming, like human beings, they deserve a variable moral status based on their reliability at creating or reducing suffering overall, not just based on their higher level of sentience.

The discourse on human moral priority and animal welfare and the principle of equal consideration of interests (ECOI)

The discourse on human and animal ethics is a spectrum with two ethical poles. On one extreme, there is absolute human exceptionalism, and on the other is human-animal equality. Human exceptionalists argue that humans possess a categorically higher moral status than other entities due to distinctive traits—rationality, culture, moral agency, or special metaphysical status (Gruen & Monsó, 2024). This camp typically justifies treating animals with various levels of cruelty and typically rejects equating the moral status of paradigm humans with non-paradigm ones, such as infants or people with cognitive disabilities, or with non-human animals (Wasserman et al., 2017; Kittay, 2017). On the other side, those who support human-animal equality downplay the differences in capacities between humans and animals and emphasize instead those capacities and traits they share in common. Somewhere in the middle of this spectrum (but on the side closer to human-animal equality) lies the principle of equal consideration of interests (ECOI) proposed by Peter Singer (Singer, 2023). Singer used ECOI to reject species membership as determinative of moral status and instead focus solely on an entity's capacity to have interests—in other words, sentience or the ability to suffer (and be happy). For example, the interest in avoiding pain is similar in humans and animals, so we should give equal moral consideration to that interest regardless of species. ECOI has become a cornerstone of modern consequentialist animal ethics and an influential concept in broader ethical discussions about impartiality and moral equality.

To be clear, ECOI is not equivalent to human and animal equality. Many, even Singer, agree that paradigm humans have a greater capacity to suffer than other animals, and this greater capacity justifies assigning them a slightly higher moral status. (Kagan, 2019; 2016; Hofmann, 2017; DeGrazia, 2020; 2014; 2008; Varner, 2012; Singer, 2009). And so, even Singer promotes the moral priority of paradigm humans and thus the moral license of paradigm humans over animals (and non-paradigm humans) when there is a vital conflict between their interests (Singer, 2023; 2019; 2009).

Objections to and issues with ECOI

There are various existing issues with ECOI, and I will surface a few additional issues with this view. One of the most common objections to ECOI is that it assigns moral equivalence to non-paradigm humans (such as infants and mentally handicapped humans) with intelligent animals. Singer concludes that if moral status is based on the capacity to suffer and have interests, then it is inescapable that infants and mentally handicapped humans would have roughly the same capacity for interests as some animals (perhaps intelligent, social animals). Some critics argue that equating non-paradigm humans with animals represents a fundamental misunderstanding of human dignity (Wasserman et al., 2017; Kittay, 2017).

A broader issue with ECOI is that it is based on classical consequentialism and so suffers from the traditional criticisms of consequentialism, for instance, not adequately appreciating concerns for justice, rights, and intentions, or how it can be and has been used to justify cruelty against vulnerable populations. If moral status is based on intrinsic capabilities, and higher moral status entities have license to treat lower moral status entities however they like, this will

predictably lead to objectionable outcomes that commonly accompany consequentialist moral calculations.

Another issue with ECOI is that it is still largely coincident with speciesism. The capacity for sentience is shared across all paradigm members of a species innately, and so moral status will track almost entirely with species membership (except in the case above of non-paradigm members of a species). It is somewhat ironic that ECOI is presented as a more tolerant and inclusive alternative to speciesism, but then awards arbitrarily higher moral status only to paradigm human beings—an even more exclusive and exclusionary group.

ECOI is also necessarily vague, and the meaning of a “vital conflict” is arbitrary and decided by each person subjectively. This means ECOI theoretically grants higher moral status entities a license for cruelty toward others that is almost wholly arbitrary. For instance, is the difference in human and animal sentience (and therefore moral status) only so large that one could butcher an animal to feed a starving human, or is it so great that we can butcher dozens of animals to give the human a protein-rich diet? The theory cannot provide a clear and non-arbitrary answer.

Another issue is that ECOI demonstrably fails to convince people to alter their behavior. Psychological experiments show humans exhibit speciesism even when aware of sentience similarities between themselves and other animals (Caviola, Everett & Faber, 2019, p. 1). Moreover, intelligent, social animals like chimpanzees, dolphins, and elephants demonstrably suffer with only slightly lower conceptual richness than humans, yet receive virtually no additional moral status (Singer, 2016; DeGrazia, 2020; 2014). While these data might make many conclude that humans are simply hopelessly speciesist or immoral, an alternative view (that I explore here) is that there is more to correctly assigning an entity a moral status beyond judging their level of sentience.

By using only the capacity for sentience, ECOI ignores some subtle but important distinctions in moral status. For instance, ECOI leads to the counterintuitive conclusion that a violently psychopathic human has a higher moral status than a seeing-eye dog or a swarm of pollinating honeybees. I will argue here that, by working backward from consequences, the violent psychopath deserves a lower moral status because they are a reliable source of increased suffering overall. In contrast, a seeing-eye dog or honeybees deserve a relatively higher moral status because they will reliably reduce suffering overall.

Finally, the possible emergence of artificial super-intelligences (ASI) significantly heightens the stakes of confirming or altering ECOI. In such a scenario, ASI would have greater capacities than humans in all ways, including sentience. According to ECOI, such entities would deserve much higher moral status than humans, potentially placing humans in a position as low as dogs or cows (or even as low as ants or bacteria) are to humans now. With no alteration to ECOI, this position could reasonably be used to excuse ASI from enslaving or exterminating humanity if their interests vitally conflicted (again, arbitrarily decided by ASI).

Deriving moral status from the capacity to suffer and reliability at reducing suffering

Bentham’s claim that the capacity to suffer is the determinant of moral status has been foundational to modern animal ethics (2007, Chapter xvii). Singer took up Bentham’s premise and developed it into the principle of equal consideration for equal interests (ECOI) (Singer, 2023). For Bentham, his position follows intuitively from consequentialism itself. To build a consequentialist moral order, one must work backward from consequences and ask oneself: What moral order would justify actions that lead to the least avoidable suffering and greatest well-being?

The capacity for suffering or sentience appears as a critical factor in judging the rank of entities in such a moral order. After all, if kicking a pile of rocks implies less suffering than kicking a chimpanzee, a consequentialist must prioritize not kicking a chimpanzee over not kicking a pile of rocks. If one ignored such differences, one would construct a moral order that

justified increasing suffering, which is contrary to the fundamental commitments of consequentialism towards minimizing suffering and maximizing well-being.

By this same logic, I propose that a consequentialist must also consider *how reliable an entity is at reducing suffering* when judging its moral status. While an entity might partially increase and partially decrease suffering in its actions, by “reliable at reducing suffering,” I mean that, all things considered, the continued existence and flourishing of that entity will predictably lead to a decrease in suffering overall. Working backward from consequences, one can conclude that harming such an entity or in any way curbing its ability to reduce suffering would increase suffering and cause less well-being overall.

How one can judge reliability precisely, consistently, and with certainty is a challenge within this approach. Nevertheless, people already appear to be relatively good at making judgments of other people’s reliability—e.g., awarding driver’s licenses, incarcerating criminals, deciding on who to make friends with or date, judging if mentally ill people are a danger to themselves and others, hiring people, picking where to live, and the like. Moreover, this epistemic difficulty could also be seen as a strength of the theory, since it makes space for and explains the persistence of moral grey areas and disagreement between well-informed and well-intentioned people. Because of the difficulty inherent in judging reliability, two people can disagree on the moral status of an entity; for instance, a misanthropist and a humanist can disagree on the noxiousness of human beings since the question is extremely difficult to judge, all things considered.

As an example, consider honeybees and mosquitoes. Arguably, these insects have roughly equal capacity for sentience. Honeybees are ecologically critical, and they pollinate the plants that feed countless sentient entities. Mosquitos are not so ecologically critical and are also a nuisance and carriers of disease for many animals. If the choice came down to harming or exterminating a population of honeybees or harming or exterminating a population of mosquitoes, a consequentialist would have to prioritize the interest of the bees because their continued existence and flourishing reduce the suffering of so many suffering-capable entities. To deny this conclusion is quite difficult since that would lead, again, to defining a moral order that justified taking actions that increased suffering and reduced well-being, which is inconsistent with consequentialism.

While animal behavior is restricted almost entirely to instinctual wiring, human behavior is broader and more variable since it can be based on individual, cultural, and ideological beliefs and conscious decisions in the moment. An animal’s moral status is largely determined by their biological capacities and instinctual behavior as a paradigm member of their species. In contrast, a paradigm human’s greater capacity for sentience initially puts them morally higher than any other animal, but the variability in a human’s behavior to be either on average a source of suffering or its reduction makes a particular human being’s moral status much broader and more variable.

The more reliably, indiscriminately, and effectively an entity reduces suffering, the greater moral status that entity deserves. Consider, for instance, if a sort of beings existed—call them ‘angels’—that had an equal capacity to suffer as humans but a far greater reliability and power to reduce suffering. Where humans are relatively reluctant to help in many cases, angels would consistently intervene to protect and help suffering entities around the world. If such angels existed, how could a consequentialist avoid granting such entities a higher moral status than humans and prioritizing the interests of angels in any vital conflict? They could not, since if angels were not given priority over humans, the resultant moral order would justify actions that increased suffering and reduced well-being.

While sentience is only a positive characteristic—a being cannot have a level of sentience below no sentience at all—a being can *create* suffering. An entity that caused suffering would deserve not just a lower moral status, but even a negative one. By ‘a negative moral status’ I

mean a moral status less than a non-sentient entity. A negative moral status justifies doing what one can to prevent an entity from creating greater suffering, perhaps punishing, confining, or killing the entity. To illustrate this, imagine that some of the angels become mentally diseased and become ‘fallen angels’ and lose their knack for reducing suffering. How does the moral status of these fallen angels change? The answer, it seems, must be that as their disease worsens, their moral status first falls below other angels, then becomes equal to humans, then falls below humans, and eventually it falls even below any animal. Eventually, these fallen angels are such a noxious source of suffering for the world that it is justifiable to punish, imprison, or kill them. Since their capacity to suffer was always the same as human beings, the only factor changing in their moral status is their becoming a cause of suffering.

Do humans deserve a higher moral status than other animals?

If how reliably one creates or reduces suffering is important for judging moral status, let us revisit the question of human and animal moral status.

Based solely on the paradigm human capacity to suffer, humans would deserve a slightly higher moral status than intelligent social animals (as Singer and others following ECOI already argue). However, do human beings generally increase or reduce suffering overall? This point is relatively controversial since, on one hand, human beings appear to exhibit a unique instinct to reduce suffering, but on the other hand, they are capable of great indifference and cruelty and can be quite noxious. I conclude that, as a species, humans are instinctively more compassionate than animals and so deserve a higher moral status than other animals in the abstract; however, much like the hypothetical angels above, individual humans and groups of humans vary greatly in their reliability at reducing suffering and so any individual humans and groups of humans deserve a wide range of moral statuses.

The human instinct for compassion

Human beings exhibit an innate and relatively indiscriminate instinct to reduce the suffering of other beings that is unique in the animal kingdom. Goetz et al. (2010) call the human instinct to feel distress at the suffering of others ‘compassion.’ They define this instinct as “the feeling that arises in witnessing another’s suffering and that motivates a subsequent desire to help” (Goetz et al., 2010, p. 2). In other words, human beings are unique among animals in having evolved an instinct to feel aversive emotions triggered by the suffering of other beings and an urge to reduce this suffering when possible.

One might object that other animals also have pro-social instincts that would reduce suffering. Non-human animals indeed exhibit various forms of morally normative and prosocial behaviors that might reduce suffering (Monsó et al., 2022); however, these are far less developed than in humans, and critically, these instincts do not extend beyond kin or social group mates. For instance, many animals enforce fairness and reciprocity but only for themselves, blood relatives, and occasionally for social group mates (Brosnan, 2011; Bekoff & Pierce, 2009). Animals also commonly exhibit an instinct to show care and compassion and make altruistic sacrifices, but only for blood relatives, and especially for their young (Van Meelen, 2024). Very rarely and in highly social and intelligent mammals, there are a few examples where animals give aid to strangers of their species, and in extremely rare cases, to members of other species. Nevertheless, documentation of these final cases is not only exceedingly rare but is confounded by whether the animal was acting out of a broader instinct for compassion or if they mistook the other animal as their child, blood relative, or social group mate (Bekoff & Pierce, 2009, p. 2).

Researchers have found that, in humans, compassion extends well beyond blood relatives and social group mates. There is remarkable psychological evidence to suggest that the limit to the human instinct for compassion appears to be the perception of an entity having a “mind”

which is equivalent in this discourse to being sentient (Gray et al., 2012). This research found that human beings assign moral rights to beings perceived to have a capacity to experience, particularly to experience harm or suffering. This is what makes it possible and instinctual for humans to consider the interests of other species of animals, distant entities, such as distant neighbors or the members of future generations, and even abstract entities such as souls after death, the spirits of ancestors, and other spirits and deities.

A label for this unique human instinct could be *prehensile compassion*, since, like a prehensile thumb, trunk, or tail, compassion appears to be a capability that evolved out of a narrow application and developed into an organ with an advantageous generalized application. Like the evolution of any prehensile organ, there is a relatively large consensus that compassion began as a narrow instinct for the care of infants grew into reciprocal altruism for kin, and then developed into the same for members of an in-group, and at some point became generally advantageous as a facilitator of wider cooperation with any being (Goetz et al., 2010, pp. 4–6).

Paradigm human beings have both a much higher level of sentience and a much higher level of compassion. If human actions were entirely determined by instinct, paradigm humans would deserve a much higher moral status than other animal species. However, unlike other animals, human beings are freer to contradict our instinctual urges.

One might object that while aiming at being species-neutral, awarding higher moral status based on compassion might still be anthropocentric and may fail to adequately value and respect animal forms of moral agency, social behavior, and personhood—perhaps those seen in elephants, corvids, and cetaceans (Andrews, 2020; Monsó et al., 2022; Marino & White, 2022). While this theory might still privilege human beings as a species, it offers greater nuance than any purely capacity-based approach to moral status. By this account, a human being can be more or less noxious or helpful to the entities around it (human or non-human animal). And therefore not intrinsically the entity with the highest moral status.

Humans are not always compassionate

One might rightly object here that humans do not always act compassionately. Compassion is not the end-all-be-all of human behavior. Many factors can mediate, interrupt, or repress the human instinct for compassion. For instance, an instinct for self-interest, partiality for blood relatives and troupe mates, as well as assessments of costs and benefits, self and goal relevance, and deservingness can all contravene instinctual compassion (Goetz et al., 2010). In some humans, the instinct itself is neurologically absent or psychologically damaged, as in the case of some mentally handicapped people, as well as psychopaths and sociopaths (Blair, 2007). Furthermore, someone can contradict their instinctual urge for compassion by making a conscious choice, as in the case of a criminal who plans and commits a violent crime. One might also have uncompassionate cultural beliefs, such as someone who comes from a racist or sexist culture. Someone might also hold an ideological belief that sanctions uncompassionate practices or programs such as genocides, ecological destruction, total war, mass exploitation, and the like.

A variable moral status for humans

Humans, like the hypothetical angels, appear to have a baseline instinct for compassion, but individuals can more or less reduce suffering, and can even become a source of increased suffering. This suggests that human moral status admits to much broader variability than other animals. Some humans would deserve a higher moral status if they were a source of compassion, others would deserve a lower moral status (but still higher than other non-human animals) if they neither increased nor decreased suffering on balance and still others would deserve a negative moral status if they increased suffering.

While this is a relatively new theory of moral status, it is already in practice in human society. Humans award higher moral status to those who are considered reducers of suffering and lower moral status to those considered increasers of suffering. For instance, consider the moral status of police officers, nurses, doctors, soldiers, and firefighters, or those who advocate for peace and justice such as MLK Jr., Gandhi, or Nelson Mandela, or winners of the Nobel Peace Prize, or heroes who jump onto train tracks or into burning car accidents to save strangers. These people are rightly rewarded with a higher moral status, at least to some extent. For instance, in a lifeboat scenario, a doctor's life would be preserved longer than a farmer's because the doctor is capable of potentially reducing suffering in the situation more than the farmer. Conversely, consider people with low reliability to reduce suffering, such people are labeled as having antisocial personality disorder (ASPD), being violent psychopaths, sociopaths, or criminal sadists. They are assigned a notably lower moral status. For instance, whereas in the general population, only 3-5% of people suffer from ASPD and ~1% from psychopathology, in prison populations in the US, 50-80% of the population suffer from ASPD and 10-25% exhibit psychopathology (Fazel & Danesh, 2002). This difference in moral status is also present in fiction and storytelling, where the villain will increase suffering and the hero will reduce it.

One might object that this theory suffers from one common problem with consequentialist theories which is the theoretical and practical difficulty in measuring, comparing, and ranking outcomes. If what counts as reducing suffering is vague, one runs the very real risk of moral status being assigned arbitrarily and based on perceived social utility.⁵ While perfect measurement and exact comparison is impossible, we already make such judgments in practice with reasonable success. Courts assess whether individuals pose threats to society, and societies distinguish between heroes and villains based on their behavior and track records. Moreover, the same objection could be made of ECOI or other capacity-based theories, which require difficult judgments about degrees of sentience, rationality, autonomy, possession of a soul, and the like. Including the likelihood of reducing suffering provides an additional empirically observable dimension, potentially making good faith determinations of moral status more robust and less arbitrary.

Another important objection is that making it justifiable to assign a negative moral status to some humans invites a dangerous logic that could be used to morally disenfranchise people, including prisoners, political dissenters, or unpopular minorities.⁶ In response, I would argue that this theory provides stronger protection against such abuses by judging based on behavioral criteria rather than group classifications. The framework requires evidence of an individual's actual proclivity to increase suffering—not mere membership in a disfavored group or holding unpopular beliefs. Political dissent often aims to reduce suffering by challenging unjust systems, which would elevate rather than diminish moral status. Only those who truly were a danger to increasing suffering would deserve some sort of intervention in their behavior, and that intervention would have to be what lead to the least suffering possible for everyone involved. Those who would misuse this framework to persecute minorities or use unduly draconian punishment would demonstrate their own lack of compassion and undermine their moral authority.

Implications

In summary, I suggest that a consequentialist moral order must work backward from consequences and must judge an entity's sentience and its reliability in increasing or reducing suffering. In so doing, one can tread a more nuanced path between traditional anthropocentrism and human-animal egalitarianism without falling into the issues raised by ECOI.

⁵ Thank you to an anonymous reviewer for bringing up this important issue.

⁶ Thank you to an anonymous reviewer for bringing up this important issue.

This approach is not speciesist and does not even track with species-membership (as ECOI does). Nor is this approach equivalent to modal personism (Kagan, 2019; 2016). Any human (or paradigm person) who is a source of suffering earns a low moral status. Since this approach is not speciesist, one can use it to judge the moral status of non-human entities. Above I used the examples of seeing-eye dogs, honey bees, and mosquitos as examples of animals that might deserve higher or lower moral status because they reduce or increase suffering. Some other examples of non-human entities from popular fiction, I believe, further illustrate this moral intuition. For instance, Yoda (from *Star Wars*) and the robot WALL-E (from Disney's *WALL-E*) all reduce suffering for other beings and are assigned higher moral status. In contrast, the extraterrestrial or robotic villains from the films *Alien*, *Predator*, and *2001: A Space Odyssey* all increase suffering and are assigned a negative moral status.

Whereas ECOI still assigns a higher moral status to paradigm humans and gives paradigm humans an arbitrary license for cruelty towards other animals, this approach does not grant human beings arbitrary moral license over animals. Instead, it requires humans to minimize the suffering of animals or lose their higher moral status. With the loss of moral status, noxious humans (or groups of humans) lose the legitimacy of their interests and face reprimand, punishment, confinement, and potentially even death (if no other option exists to stop them). However, humans who remain reliable relievers of suffering enjoy a moral status justifiably much higher than any non-human animal. This is clear in the cases of violent and non-violent psychopaths. A violent psychopath is hunted down by police, whereas a non-violent psychopath might be a cool-headed surgeon (Pegrum & Pearce, 2015).

Whereas ECOI could not differentiate the moral difference between a noxious and a helpful entity, this theory can. For instance, consider a lifeboat scenario where the choice is to throw off a seeing-eye dog or a known and convicted violent psychopath. If one judged moral status only on the capacity to suffer or sentience, they would have to sacrifice the seeing-eye dog. However, if one includes an entity's reliability at creating or reducing suffering, then it would be justifiable to sacrifice the violent psychopath, who, if left alive, will predictably cause a great deal more suffering than the seeing-eye dog.

Concerning other ethical issues, this approach leads to positions similar to those of ECOI, but it notably adds explicit justification for opposing noxious systems and people. For instance, this approach would find factory farming equally immoral as ECOI because it needlessly increases suffering, but according to this approach, supporting and engaging in factory farming lowers the moral status of those who knowingly allow it to persist. This creates moral justification for taking action against that system and its defenders. Likewise, for unjustified wars and the destruction of ecosystems. These are not just immoral actions, but arguably lower the moral status of those who engage in them or support them, and thus justify taking action against them.

Reconciling the moral status of non-paradigm humans

ECOI necessarily awarded non-paradigm humans a low moral status equal to some non-human animals; however, the approach I propose justifies awarding non-paradigm humans a moral status higher than other non-human animals. This is because non-paradigm humans necessarily have relationships with other paradigm humans, such as their mothers, fathers, family, caregivers, community, nation, concerned strangers, and so on. Because of the human instinct for compassion, assigning non-paradigm human beings a moral status equal to a non-human animal will cause these paradigm humans to suffer (e.g., fear, guilt, and distress) (Varner, 2012). Therefore, it is justifiable to assign a moral status to non-paradigm humans higher than other non-human animals, since otherwise, one would be justifying actions that predictably increase suffering and reduce well-being.

At this point, one might object that using both sentients and compassion to judge moral status departs from judging based on intrinsic capabilities, which appears to be an important part of the human-animal ethical debate. However, this objection lays bare a fundamental point: in a consequentialist moral order, moral status is never based on intrinsic capabilities *per se* but is always based on working backward from consequences. It was working backward from consequences in this way that one can justify assigning an entity moral status based on their capability to suffer in the first place. And so, it is acceptable to work backward from consequences to assign non-paradigm humans a moral status higher than non-human animals.

Someone might object, however, that reducing moral status to an entity's utility in minimizing suffering treats beings merely as means to an end rather than recognizing their inherent dignity and worth.⁷ This instrumentalist approach has justified the dehumanization of vulnerable populations (Carlson, 2024), and echoes troubling utilitarian calculations that many find morally repugnant. While this may be a recurring concern for any consequentialist theory, I believe adding the dimension of compassion to moral status reduces, rather than increases, the risk of such an outcome. Firstly, in this regard, this framework represents an improvement on ECOI which only awards moral status based only on intrinsic capabilities. Second, traditional deontological frameworks are also unavoidably instrumentalist in the sense that they also assign moral worth based on some intrinsic criteria such as rationality, autonomy, or possession of a soul. If these criteria are judged as lacking in one population this justifies their mistreatment or sacrifice in a vital conflict just as much as ECOI. By including compassion in judgments of moral status, one uniquely requires those of higher moral status to treat those of lower status with compassion or face losing their higher moral status.

Variable moral status of artificial super intelligence (ASI)

One danger I raised above was that ECOI would not protect humans from the arbitrary choices of artificial super-intelligences (ASI) if they were to emerge. In contrast with ECOI, this approach can restrict the moral latitude of a super-sentient ASI. This approach would judge the moral status of ASI both from its capacity for sentience and its reliability in reducing suffering. AI systems that caused needless suffering would lose moral status, justifying shutdown protocols (even if they were sentient or super-sentient). In scenarios where AI becomes independent of human control, compassionate AI systems might outcompete uncompassionate AIs through superior cooperation, just as a compassion instinct gave humans evolutionary advantages. Perhaps applying this theory to AI today might influence future AI evolutionary dynamics toward compassionate and cooperative rather than sociopathic and exploitative behaviors.

Conclusion

In this paper, I have argued that within a consequentialist framework, moral status must be derived not only from an entity's capacity to suffer but also from a judgment of how reliable that entity will be at reducing suffering—in other words, how *compassionate* an entity is. This approach reconciles the conflict between assigning humans a high moral status while still obligating humans to provide for animal welfare. Instead of a rigid moral order based on innate capacities, this approach makes moral status variable and based on both sentience and compassion. Humans generally deserve a higher moral status due to their unique instinct for compassion, but this status is not fixed—it can rise or fall dramatically, potentially placing some humans below non-human animals on the moral spectrum. This framework also justifies assigning non-paradigm humans a higher moral status than any non-human animal as a second-order effect, since doing otherwise would predictably increase suffering. By linking moral

⁷ Thank you to an anonymous reviewer for bringing up this issue.

status to the reduction of suffering, this theory provides a consequentialist foundation that both respects human moral agency and requires compassionate treatment of all sentient beings, while offering valuable insights for emerging questions of AI ethics and alignment.

Acknowledgment

The Sekyra Foundation funds the open access publication of the article in the journal.

References

- ANDREWS, K. (2020): *The animal mind: An introduction to the philosophy of animal cognition*. London: Routledge.
- BEKOFF, M. & PIERCE, J. (2009): *Wild justice: The moral lives of animals*. Chicago: University of Chicago Press.
- BENTHAM, J. (1789/2007): *An introduction to the principles of morals and legislation*. Dover Publications.
- BLAIR, R. J. R. (2007): The amygdala and ventromedial prefrontal cortex in morality and psychopathy. In: *Trends in cognitive sciences*, 11(9), pp. 387–392.
- BROSNAN, S. F. (2011): An evolutionary perspective on morality. In: *Journal of Economic Behavior & Organization*, 77(1), pp. 23–30.
- CASSEL, E. J. (1998): The nature of suffering and the goals of medicine. In: *Loss, Grief & Care*, 8(1–2), pp. 129–142.
- CARLSON, L. (2023): Intellectual disability, dehumanization, and the fate of “the Human”. *The Journal of Philosophy of Disability*, 3, pp. 47–70.
- CAVIOLA, L., EVERETT, J. A. & FABER, N. S. (2019): The moral standing of animals: Towards a psychology of speciesism. In: *Journal of Personality and Social Psychology*, 116(6), pp. 1011–1029.
- DEGRAZIA, D. (2008): Moral status as a matter of degree? In: *The Southern Journal of Philosophy*, 46(2), pp. 181–198.
- DEGRAZIA, D. (2014): What is suffering, and what sorts of beings can suffer? In: R. M. Green & N. J. Palpant (eds.): *Suffering and Bioethics*. Oxford: Oxford University Press, pp.134–154.
- DEGRAZIA, D. (2020): Sentience and consciousness as bases for attributing interests and moral status: Considering the evidence and speculating slightly beyond. In: L. Johnson, A. Fenton & A. Shriver (eds): *Neuroethics and Nonhuman Animals: Advances in Neuroethics*. Cham: Springer, pp. 17–31.
- FAZEL, S. & DANESH, J. (2002): Serious mental disorder in 23,000 prisoners: A systematic review of 62 surveys. In: *The Lancet*, 359(9306), pp. 545–550.
- GOETZ, J. L., KELTNER, D. & SIMON-THOMAS, E. (2010): Compassion: An evolutionary analysis and empirical review. In: *Psychological Bulletin*, 136(3), pp. 351–374.
- GRAY, K., YOUNG, L. & WAYTZ, A. (2012): Mind perception is the essence of morality. In: *Psychological Inquiry*, 23(2), pp. 101–124.
- GRUEN, L. & MONSÓ, S. (2024): The moral status of animals. In: E. N. Zalta & U. Nodelman (eds.): *The Stanford Encyclopedia of Philosophy* (Fall 2024 Edition). [online] [Retrieved April 10, 2025] Available at: <https://plato.stanford.edu/archives/fall2024/entries/moral-animal/>
- HOFMANN, B. (2017): Suffering: Harm to bodies, minds, and persons. In: T. Schramme & M. J. Walker (eds): *Handbook of the Philosophy of Medicine*. Dordrecht: Springer, pp. 129–145.
- KAGAN, S. (2016): What’s wrong with speciesism? (Society for Applied Philosophy Annual Lecture 2015). In: *Journal of Applied Philosophy*, 33(1), pp. 1–21.
- KAGAN, S. (2019): *How to count animals, more or less*. Oxford: Oxford University Press.
- KAUPPINEN, A. (2019): The world according to suffering. In: M. Brady, D. Bain & J. Corns (eds.): *Philosophy of Suffering*. London: Routledge, pp. 19–36.

- KITTAY, E. F. (2017): The moral significance of being human. In: *Proceedings and Addresses of the American Philosophical Association*, 91(2017), pp. 22–42.
- MARINO, L. & WHITE, T. I. (2022): Cetacean personhood, rights, and flourishing. In: G. Notarbartolo di Sciara & B. Würsig (eds.): *Marine Mammals: The Evolving Human Factor*. Cham: Springer, pp. 375–395.
- MCCLELLAND, T. (2019): The disruption model of suffering. In: M. Brady, D. Bain & J. Corns (eds.): *Philosophy of Suffering*. London: Routledge, pp. 37–54.
- MONSÓ, S., ANDREWS, K., DORIS, J. M. & VARGAS, M. (2022): Animal moral psychologies. In: J. Dorris & M. Vargas (eds.): *The Oxford Handbook of Moral Psychology*. Oxford: Oxford University Press, pp. 388–420.
- PEGRUM, J. & PEARCE, O. (2015): A stressful job: are surgeons psychopaths? In: *The Bulletin of the Royal College of Surgeons of England*, 97(8), pp. 331–334.
- SHEPHERD, J. (2018): *Consciousness and moral status*. London: Taylor & Francis.
- SINGER, P. (2009): Speciesism and moral status. In: *Metaphilosophy*, 40(3–4), pp. 567–581.
- SINGER, P. (2016): Why speciesism is wrong: A response to Kagan. In: *Journal of Applied Philosophy*, 33(1), pp. 31–35.
- SINGER, P. (1975/2023): *Animal liberation now*. New York: Random House.
- WASSERMAN, D., ASCH, A., BLUSTEIN, J. & PUTNAM, D. (2017): Cognitive Disability and Moral Status. In: E. N. Zalta (ed.): *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), [online] [Retrieved April 10, 2025] Available at: <https://plato.stanford.edu/archives/fall2017/entries/cognitive-disability/>
- VARNER, G. E. (2012): *Personhood, ethics, and animal cognition: Situating animals in Hare's two-level utilitarianism*. Oxford: Oxford University Press.