Kent Academic Repository

Downloaded from

The version of record is available from

This document version

DOI for this version

Licence for this version

Additional information

Versions of research works

Versions of Record

Author Accepted Manuscripts

Enquiries

Utilitarianism for animals, Kantianism for people? Harming animals and

humans for the greater good

Forthcoming in the *Journal of Experimental Psychology: General*

Lucius Caviola[1,2], Guy Kahane[3], Jim A.C. Everett[3,4], Elliot Teperman, Julian

Savulescu[3], Nadira S. Faber[2,3,5]

[1]Department of Psychology, Harvard University

[2]Department of Experimental Psychology, University of Oxford

[3]Uehiro Centre for Practical Ethics, University of Oxford

[4]Department of Psychology, University of Kent

[5]College of Life and Environmental Sciences, University of Exeter

Abstract

Most people hold that it is wrong to sacrifice some humans to save a greater number of humans. Do people also think that it is wrong to sacrifice some animals to save a greater number of animals, or do they answer such questions about harm to animals by engaging in a utilitarian cost-benefit calculation? Across 10 studies ($N = 4,662$), using hypothetical and real-life sacrificial moral dilemmas, we found that participants considered it more permissible to harm a few animals to save a greater number of animals than to harm a few humans to save a greater number of humans. This was explained by a reduced general aversion to harm animals compared to humans, which was partly driven by participants perceiving animals to suffer less and to have lower cognitive capacity than humans. However, the effect persisted even in cases where animals were described as having greater suffering capacity and greater cognitive capacity than some humans, and even when participants felt more socially connected to animals than to humans. The reduced aversion to harming animals was thus also partly due to *speciesism*—the tendency to ascribe lower moral value to animals due to their species-membership alone. In sum, our studies show that deontological constraints against instrumental harm are not absolute but get weaker the less people morally value the respective entity. These constraints are strongest for humans, followed by dogs, chimpanzees, pigs, and finally inanimate objects.

Keywords: moral judgments, utilitarianism, harm aversion, speciesism, human–animal relations

Utilitarianism for animals, Kantianism for people? Harming animals and humans for the greater good

Utilitarianism tells us to always maximize the good. This is an attractive moral aim in many cases—few would deny that, for example, we should try to save as many lives as we can, and generally minimize harm to others. In some cases, however, most people think that it would be wrong to maximize the good because this would violate what moral philosophers call 'deontological side constraints' (Nozick, 1974)—moral rules that forbid harming, torturing, or killing innocents, even for a good end. A great deal of recent psychological research has examined whether and under what circumstances people hold deontological constraints (e.g., Greene, 2014). For example, in the famous Footbridge trolley case (Thomson, 1984) a large majority of people are reluctant to push one person off a bridge to save the lives of five people trapped on the tracks of an oncoming train (Greene, 2009), and it has even been claimed that such constraints are universal and innate (Mikhail, 2007). What has not been studied yet is whether people also apply deontological rules to non-human animals (from now on simply 'animals'). Are people as reluctant to harm animals to save a greater number of animals as they are for humans? And if indeed people do accept such constraints on harm to animals, are these deontological constraints similar in structure and strength to those that protect humans?

Kant (1785) offered one answer to these questions. He argued that the source of the deontological constraints that protect us from being sacrificed for the sake of others is our unique rational nature—our rationality makes us priceless and, therefore, we cannot be an input into a utilitarian cost-benefit calculation. On Kant's own view, this meant that, because animals are not rational, they do not morally matter in their own right, though he frowned upon gratuitous cruelty to animals because it could make us more willing to harm humans. The philosopher Robert Nozick (1974) famously discussed an alternative principle he

referred to as: "utilitarianism for animals, Kantianism for people". On this view, which

Nozick himself did not entirely endorse, animals do matter, but they matter very differently

than humans. As non-rational individuals, they do not enjoy deontological protections and

can be entered into utilitarian calculations. The philosopher Judith Jarvis Thomson similarly

writes that, unlike humans, "animals do not have claims to not be killed… Would it be

permissible to kill one chicken to save five chicken? I think it would." (Thomson, 1990).

It is clear that people value animals less than humans. This universal phenomenon

shows across various contexts and has been investigated by both philosophers (e.g., Singer,

1975) and psychologists (e.g., Amiot & Bastian, 2015; Caviola, Everett, & Faber, 2019;

Dhont, Hodson, & Leite, 2016). Psychological research has shown that to large extent people

morally prioritize humans over animals simply on the basis of species-membership—a

tendency referred to as *speciesism* (Caviola et al., 2019; Horta, 2010; Singer, 1975). Still it

remains an open question how exactly people value animals (cf. Goodwin, 2015). In

particular, we are interested in what types of deontological constraints against harming

animals or humans people may hold. We focus on the context of sacrificial moral dilemmas

to study whether people are willing to harm a few animals or humans to save many animals

or humans, and whether this willingness is uniform across these decision contexts. It is

obvious that people are more willing to sacrifice animals than humans for a fixed benefit,

e.g., to save many humans. However, it is unclear (i) to what extent people are willing to

harm a few animals to save many animals, and (ii) how this tendency compares to their

willingness to harm a few humans to save many humans. Current research is compatible with

multiple answers to these questions. Are people's intuitions about such dilemmas correctly

captured by Kant, utilitarianism, Nozick, or by some other view? In this paper we

systematically test a range of possible descriptive ethical views of how people apply

deontological constraints against harm to humans and animals.

Why is this of interest? First, answering this question can shed important light on the psychology of moral judgment more generally, and of moral status specifically. Typically, moral psychology has focused on scenarios involving human characters lacking concrete identities (cf. Hester & Gray, 2019). It is likely, though, that people are sensitive to *who* those characters are. In this work, we investigate whether people take into account the species-membership of the characters at stake. Do deontological constraints pervade all moral thinking, or do they only apply in the human context? If deontological constraints apply only to humans, how does morality operate with regards to other species? Focusing on dilemmas where we can sacrifice some humans to save a greater number, Greene (2014) has suggested that people engage in either utilitarian or deontological modes of moral thinking, and that when the two compete, the intuitive, deontological mode typically prevails. If deontological constraints apply only to humans, this would mean that this dual process picture applies only in the human domain, and that moral judgment in the animal domain is the result of a single process of utilitarian cost-benefit analysis. Conversely, if deontological constraints *do* apply to animals, this would raise the psychological question of whether these deontological constraints apply in the *same* way to animals as they do to humans. Moreover, since this would show that, *contra* Kant, people do not base deontological constraints in rationality or other higher cognitive capacity, it would raise the further question of *what* is the basis of such constraints—is it mere suffering capacity (i.e., sentience), the existence or lack of a social relationship with the potential victim, or something else?

Second, how people think about the moral status of animals is of intrinsic interest. Animals are routinely harmed to benefit humans (e.g., for consumption or for medical experiments), and in some cases animals are harmed to benefit other animals. It is not yet clear whether most people nevertheless hold that there are moral constraints on when and whether such harm is permissible, and what psychological processes underlie such intuitions.

Understanding these psychological processes may have potential practical implications. For example, certain wildlife conservation policies allow the killing of predator animals like wolves in order to protect other animal species. Other cases exist in the context of livestock: Should a farmer be allowed to kill a few pigs infected by the swine flu in order to prevent many more pigs to be infected and killed? Should it be legal to experiment on a few chickens in order to benefit (or bring into existence) many more chickens? Another real-world case is the question of whether it is ethically more problematic to consume certain animals than others. For example, a cow produces about 200 times as much meat as a chicken, which means that producing the same amount of meat requires killing far fewer individuals in the case of cows compared to in the case of chickens. Depending on the relative strengths in deontological constraints against harm that people hold for these animals, sacrificing certain animals to benefit humans (or other animals) may be more problematic than sacrificing others.

When is harm towards humans and animals permissible?

In order to investigate how people think about harm to animals vs. to humans, we first need to set out the theoretical options (Table 1). Our aim is to try to identify the moral views that are most consistent with lay people's intuitions. The first two views we will mention are well-known philosophical positions that we include for completeness sake even if they are unlikely to be psychologically accurate. The remaining views are all plausible descriptions of people's intuitions and have sometimes even been described as such. Note that we are not suggesting that people *explicitly* endorse one of these moral views. Rather, we ask which moral view most closely aligns with the way people think about harm to humans vs. to animals. Why this might be the case, and what underlying psychological mechanisms cause these judgments, are further questions that we will address below.

Table 1

Moral views of when it is permissible to harm humans and animals.

| Harm: | Humans | Animals | Animals |
To save:	Humans	Animals	Humans
Utilitarianism	✓	✓	✓
Kant's view	✗	✓	✓
Cross Species Deontology	✗	✗	✗
Nozick's slogan	✗	✓	✓
Multi-level Uniform Deontology	✗	✗	✓
Multi-level Weighted Deontology	weighted	weighted	✓

Note. Utilitarianism and Cross Species Deontology consider humans and animals to have equal moral status (assuming they are equally sentient). All other views consider humans to have higher moral status than animals (even if they are equally sentient). While Kant thought that the moral status of animals is not fundamentally different from that of objects, he did think that harming animals gratuitously is wrong on instrumental grounds, because it can make us more willing to harm humans.

The first possibility is that people's intuitions are best captured by *utilitarianism*. On this account, the same harm matters equally, regardless of who suffers it (weighted by their degree of sentience)—whether humans or animals (Bentham, 1780). This view is anti-speciesist, meaning that species-membership itself should not influence the moral status of an individual (Regan & Singer, 1989). If people were strict utilitarians, they would consider it permissible (or even required) to sacrifice both humans and animals to promote the greater good (of both humans and animals). It is unlikely, however, that this view captures the intuitions of most people since, as described above, there is considerable evidence that people

accept deontological constraints against harming humans for the greater good (e.g., Greene, 2014), and people tend to value animals less than humans (Caviola et al., 2019).

Another possibility is that people's intuitions can be described by *Kant's view*. On this approach, only humans matter morally and therefore deserve deontological protection, whereas animals are just seen as objects that can be used to our own ends (Kant, 1785). The source of the value for humans is the fact that we, but not animals, are rational individuals and possess advanced cognitive capacity. This, however, is also implausible as an account of most people's views since people believe that animals do matter morally at least to some extent (Caviola, Schubert, Kahane & Faber, 2020).

Yet another possibility is that people's intuitions can be described by a view we can call *Cross Species Deontology*. On this account, the same deontological principles apply in the same way to all species. Neither humans nor animals should be sacrificed for the greater good of either. Like utilitarianism, this view is anti-speciesist because species-membership itself should make no difference (for a similar view, see Regan, 1987). Abolitionist animal rights activists tend to endorse similar moral positions (e.g., Francione, 1995). This again is unlikely to capture the common view given that people on average think that it's permissible to harm animals to benefit humans, e.g., via medical testing (Caviola et al., 2019).

There are, however, at least three ways to capture this intuitive moral difference between humans and animals while still ascribing some moral significance to animals, *contra* Kant. The first is the Nozick (1974) suggestion, discussed above, that deontology applies only to humans, while utilitarianism applies to animals. When it comes to animals, we should simply maximize utility—even by sacrificing some for the greater good. But when we turn to humans, this is forbidden. Many moral philosophers take Nozick's suggestion very seriously (cf. Thomson, 1990; McMahan, 2002; Kagan, 2019; Killoren & Streiffer, 2019) and the Nuffield Council on Bioethics has even called utilitarian cost-benefit analysis the

"cornerstone" of research on animals (Kilkenny, Browne, Cuthill, Emerson, & Altman, 2010; Nuffield Council on Bioethics, 2005). Moral philosophers often assume that Nozick's suggestion reflects people's intuitions, and that it best captures the commonsense view about animals (Kagan, 2019; Killoren & Streiffer, 2019). For example, people generally consider keeping animals in well-run zoos as morally permissible. Keeping innocent humans imprisoned, in contrast, is generally considered repugnant even if they were kept very happy. Another example is animal research. Most Western jurisdictions have a near absolute prohibition of more-than-minimal research on children who, like animals, cannot give consent (Gennet & Altavilla, 2016)—even if it could be beneficial in expectation. Regulations and intuitions concerning research on animals are much more permissive (Varner, 1994).

A second possible view is that there is a *hierarchy* of moral status (cf. Kagan, 2019). Individuals that are lower in the hierarchy (e.g. pigs) can be sacrificed for the sake of those higher up (e.g. humans)[1]. But *within* each level of moral status, the deontological constraints offer the same protections (i.e. it's wrong to sacrifice a pig to save five pigs), and these protections apply to the *same* degree. We call this *Multi-level Uniform Deontology*.

A final, more complex view, which we call *Multi-level Weighted Deontology*, is that the deontological protections are not absolute, and get *weaker* the lower the level of moral status. As we go down the hierarchy, the less stringent the deontological constraints. According to this view, people would consider harming animals to save many animals neither

---

[1] Note that some philosophers *define* moral status in terms of the deontological protections that individuals enjoy. We, in contrast, assume that moral status is a more abstract construct which *determines* the extent to which people perceive it permissible or required to harm or help a certain individual. Moreover, moral status relates to beliefs about the moral status people attribute to individuals in an absolute sense and not beliefs about having special obligations to certain individuals. For example, people might think it is justified to prioritize family members over strangers while still believing that both possess the same moral status and rights.

completely permissible nor completely wrong (as Multi-level Uniform Deontology would),

but instead somewhere in between. Further, the lower the moral status of the animal in

question, the more permissible they would consider harming it to save many animals with the

same moral status (i.e., it is more permissible to sacrifice one cow to save five others than to

sacrifice one human to save five others[2]). In cases where the moral status of a being is very

low (as for example with certain animals or objects), the implications of the Multi-level

Weighted Deontology will resemble those of Nozick's slogan because the deontological

constraints will be low or non-existent.

These three hypotheses have not yet been tested. Our own hypothesis is that Multi-

level Weighted Deontology is the view that describes people's intuitions best. A growing

body of evidence suggests that emotional aversion to harming others plays a key role in

driving deontological constraints against harming few humans to save a greater number

(Crockett, 2013; Cushman, 2013; Greene, 2014; Wiech et al., 2013). Aversion to harm is a

matter of degree, and many people clearly feel some aversion to harming animals through

direct, 'personal' acts, even if to a considerably lesser degree than in the human case. If so, it

seems unlikely that they would approach sacrificial choices involving animals in a purely

utilitarian manner, nor that the deontological inhibitions against such harm would be as

strong as in the human case. We therefore expect people to consider it permissible to harm

animals to benefit humans and that they consider it somewhat, but not completely,

permissible to harm animals to save many animals with the same moral status. Since people

---

[2] The weighted deontological constraints can show both in terms of a changing *degree* of permissibility of harming a few to save many but also in terms of a changing *threshold* of the minimum number of beings that need to be saved in order to make it permissible to harm a few. For example, people consider it more permissible to harm one cow to save five cows than to harm one human to save five humans (changing degree of permissibility), and people believe that a lower number of saved cows is required to make it permissible to kill ten cows than the number of saved humans that is required to make it permissible to kill then humans (changing threshold).

value animals much less than humans, we assumed, the deontological constraints they ascribe to them are likely weaker even in intra-species cases.

Deontological constraints are thus not absolute but can get weighted and can sometimes be overturned (Holyoak & Powell, 2016; Kahane, 2015). For example, people are likely to consider it less wrong to harm a few humans to save a *very* large number of humans (Tremoliere & Bonnefon, 2014). Similarly, we expect that they will neither have absolutely strong nor non-existent deontological constraints against harming animals even in the intra-species context but rather moderately strong constraints. Were we to find, however, that people roughly have the same deontological constraints for humans and animals in the intra-species context, or always find it entirely permissible to sacrifice few animals to save a larger number, that would falsify Multi-level Weighted Deontology.

Potential psychological mechanisms

The different moral views discussed above leave the question unanswered what the underlying psychological mechanisms are. So far, no studies have systematically explored judgments in sacrificial dilemmas that involve animals. An exception is a recent study, in which participants were presented with two mice cages attached to an electric shock machine (Bostyn, Sevenhant, & Roets, 2018). Participants were informed that, as a default, the five mice in one of the cages would receive a painful electric shock unless participants decided to intervene and push a button that would instead redirect the electrical current to the other case that contains just one mouse. 84% of participants pushed the button. In contrast, Bostyn et al. found that when the same situation was described hypothetically, only 66% said that they would push the button. However, since the study did not directly compare cases involving humans against cases involving animals, we cannot draw a clear conclusion from it. Further,

this study involved an 'impersonal' moral dilemma and multiple studies have found that most people regard such harm as permissible even in in the human case.

As mentioned above, previous research on sacrificial moral dilemmas has identified a central role for harm aversion in explaining moral judgments in sacrificial moral dilemmas (Crockett, 2013; Cushman, 2013; Greene, 2014). The more aversive people are to harm the few, the less willing they are to harm them for the greater good. When harm aversion is reduced, such as in cases where the harm is less personal (e.g., when one has to hit a switch to harm), people are more willing to harm the few for the greater good, than if the harm is more personal (e.g., when one has to physically push someone). There is, therefore, considerable evidence that, in the human case, the judgments that mirror deontological constraints reflect aversion to harm. Thus, it is plausible that this is also the case when we turn to animals.

Therefore, one possible explanation for why people might be more willing to harm a few animals to save more animals than to harm a few humans to save more humans could be that they have a weaker aversion to harm animals in general. Studies have shown that people perceive humans to be more deserving of prevention of harm than animals (Caviola et al., 2019; Everett, Caviola, Savulescu, & Faber, 2018). Further, people also endorse harming animals to save humans (Topolski, Weaver, Martin, & McCoy, 2013; Awad et al., 2018; Petrinovich, O'neill, & Jorgensen, 1993; Topolski et al., 2013).

If reduced harm aversion for animals makes such harm seem more morally permissible, this raises the further question of *why* harm aversion for animals is weaker. One possibility is that people are more willing to harm animals than humans because they perceive animals to suffer less than humans and to have a lower cognitive capacity than humans. Both philosophers as well as psychologists (H. M. Gray, Gray, & Wegner, 2007; K. Gray, Young, & Waytz, 2012) have emphasized the importance of suffering capacity (or

sentience) and cognitive capacity (or intelligence / rationality) for moral status attribution. Indeed, it has been shown that people tend to de-mentalize animals, i.e., they attribute reduced mental capacities (both suffering capacity and cognitive capacity) to animals (Bastian, Loughnan, Haslam, & Radke, 2012). And the lower the perception of suffering capacity of animals is, the less people value the animals (Caviola et al., 2019). A recent study found that in trolley-like scenarios people were more likely to harm robots than humans to save many humans (Nijssen, Müller, Baaren, & Paulus, 2019). Yet the more human-like the robots were, the less likely people were to harm them, which was explained primarily due to the attribution of affective states. Similarly, in previous work we have shown that the perceived cognitive capacity can play a small though significant role in explaining the tendency to prioritize humans over animals (Caviola et al., 2020). For example, it seems plausible that people would consider harming animals with a low cognitive capacity (e.g., birds) more permissible than harming animals with a higher cognitive capacity (e.g., apes). Together, these findings suggest that perceived suffering capacity as well as cognitive capacity may play a part in explaining why people are more willing to harm animals than humans.

A further, not mutually exclusive possibility is that people are more willing to harm animals than humans simply on the basis of their species-membership alone, regardless of the perceived capacities of the being, i.e., speciesism (Caviola, et al., 2019; Dhont et al., 2016). Indeed, previous research has demonstrated that people have speciesist attitudes. For example, they prioritize humans over animals (in cases of helping) even when perceived suffering capacity is taken into account (Caviola et al., 2020). We also found that speciesism can modulate perceived suffering capacity. That is, the more speciesist people are, the less they perceive animals to be capable of suffering (Caviola et al., 2019), which could in turn reinforce reduced harm aversion for animals. Thus, it is likely that perceived suffering

capacity, perceived cognitive capacity, and mere species-membership (speciesism) together contribute to reduced harm aversion for animals.

Thus, our hypothesized model of the psychological mechanisms, which we refer to as *Harm Aversion Mediated Speciesism*, is that people see harming animals for the greater good as more permissible than harming humans for the greater good because of a generally weaker aversion to harm animals. Harm aversion, in turn, is multiply determined: it is driven by perceptions of suffering capacity, perceptions of cognitive capacity, and speciesism. Crucially, speciesism reduces harm aversion more for animals than for humans. In short, the model that we will test is a moderated mediation from species condition (humans vs. animals) to harm aversion to degree of moral permissibility (of harming a few to save many of the same species), whereby the effect of condition to harm aversion is moderated by speciesism and where harm aversion is also predicted by perceived suffering capacity and perceived cognitive capacity. We do not rule out the possibility that there are effects from species (humans vs. animals) onto moral permissibility that do not go via harm aversion.

Finally, we make an additional prediction, which we consider less central to the core model but still worth exploring. We hypothesize that speciesism will be associated with reduced perceived mental capacities (both suffering capacity and cognitive capacity) in animals but less so in humans.
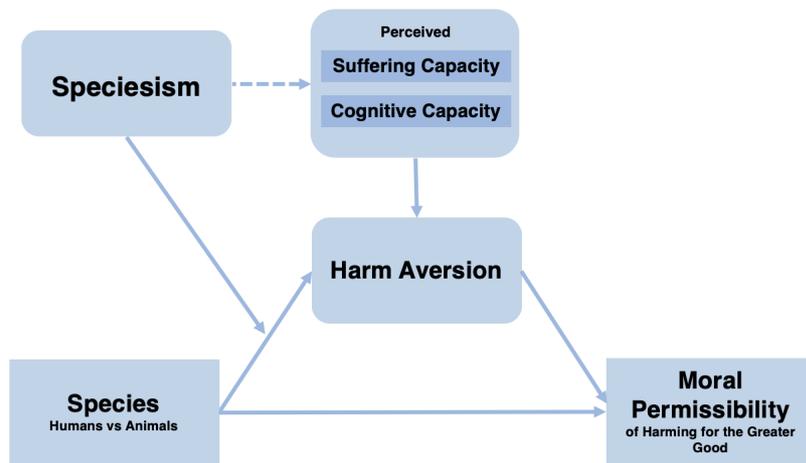
Figure 1. The Harm Aversion Mediated Speciesism Model describes the psychological mechanisms underlying judgments about the permissibility of harm to animals and humans. A moderated mediation from species (humans vs. animals) to harm aversion to moral permissibility, whereby the effect of condition to harm aversion is moderated by speciesism and where harm aversion is also predicted by perceived suffering capacity and perceived cognitive capacity.

The present research

In ten studies we tested the extent to which people accept deontological constraints against harming animals in both intra- and inter-species contexts. And we compared it to the extent to which people accept such constraints for humans. Our first aim was to test whether indeed people's judgments can best be described by Multi-level Weighted Deontology. Our second aim was to test the Harm Aversion Mediated Speciesism Model. We tested this in studies 1, 2, 3, 7, 8, 10. The cognitive capacity factor of the model was only measured and tested in Study 8. In all studies participants were presented with different (usually hypothetical) sacrificial moral dilemmas, which involved harming either animals or humans in order to prevent harm to a larger number of animals or humans and asked for participants' judgments of the moral permissibility of doing so.

Open science. As for all studies in this paper, reports of all measures, manipulations, and exclusions, and all data, analysis code, and experimental materials are available for download at: https://osf.io/nt69s/ Studies 1, 2, 5 and 9 were pre-registered on the Open Science Framework.

Ethics statement. For all studies, relevant ethical guidelines were followed, and the research was approved through University of Oxford's Central University Research Ethics Committee, with the reference numbers MS-IDREC-R56657/RE002.

Study 1: Humans vs. animals

In Study 1, we tested whether people are more willing to harm a few animals to save many animals (of the same species) than to harm a few humans to save many humans. Our reasoning was based on the Multi-level Weighted Deontology hypothesis, namely that people's deontological constraints are weighted by how much they value the respective individual. Since people value humans much more than animals, their deontological constraints should be much higher for humans than for animals respectively.

We tested a range of different animals, namely: panda bears, dogs, squirrels, chimpanzees, and pigs. Dogs and chimpanzees are interesting cases because they both typically are valued less than humans but more than pigs despite the fact that dogs and pigs are similar to each other in terms of level of intelligence or emotional capabilities (Caviola et al., 2019). Chimpanzees are the animal species that are most closely related to humans but still are valued considerably less than humans. Panda bears are interesting because they are often perceived as an especially likeable animals. Finally, squirrels are of interest because they are significantly smaller and of lower cognitive capacity than the other animals.

Our hypothesis—pre-registered at https://osf.io/w863g/—was that people would consider it more permissible to harm one animal to save many animals than to harm one

human to save many humans. We also assumed that there could be differences between the different types of animals but did not make specific hypotheses.

To test the Harm Aversion Mediated Speciesism Model (outlined in the introduction), we measured *harm aversion*, that is how averse to inflicting harm participants indicated to be for the respective beings in question, perceived *suffering capacity*, that is how much participants believed the respective beings to be capable of suffering when harmed, as well as *speciesism* (Caviola et al., 2019). For exploratory purposes, we also measured *empathic concern* (as part of the IRI; Davis, 1980), *instrumental harm*, that is permissive attitude toward instrumental harm (OUS-IH), and *impartial beneficence*, that is impartial concern for the greater good (OUS-IB). The latter two measure individual differences in utilitarian tendencies (Oxford Utilitarianism Scale, Kahane et al., 2018). Due to the exploratory nature of these three measures we will not discuss them further throughout the paper. For all studies, correlations between the dependent variable and the follow-up scales for each condition separately are reported in Table S1 in the Supplementary Materials.

Method

Development and pre-test of materials

We developed a new moral dilemma that should be perceived as more realistic than the traditional trolley problem (Bauman, McGraw, Bartels, & Warren, 2014). Therefore, in a pre-test to Study 1 (*N* = 700; reported in the Supplementary Materials) we tested five different sacrificial dilemmas. Based on the results we have decided to rely on the vaccine death dilemma in the next studies. In this dilemma the sudden outbreak of a rare virus is described, which will kill 100 pigs (or other beings). The only way to save them is to actively infect 10 healthy pigs (that otherwise would not die) to identify the vaccine that can be used to prevent the virus from killing the other 100 pigs.

Participants

We recruited 918 US American participants online via MTurk who received $0.50 payment for their participation (in line with US minimum wage for all studies). Thirteen participants were excluded for failing at least one of two attention checks, leaving a final sample of 905 people (445 females; M_{age} = 41.31, SD = 12.53). We anticipated a small to medium effect of f = .13 between animal species based on the results the previous study that relied on abstract dilemmas mentioned above. With an alpha of 0.05, and power of 0.80, our a priori power analysis showed that we required 768 participants. Hence, we aimed to recruit 900 participants to account for exclusions.

Design, materials, and procedure

This study had six between-subjects conditions to which participants were randomly allocated to six conditions: humans, pandas, dogs, squirrels, chimpanzees, pigs. After participants read the vaccine dilemma, they were asked to indicate how morally right (i.e., permissible) or wrong they thought it to be to harm ten animals (humans) to save one hundred animals (humans) on a 7-point scale from 1 (*Absolutely morally wrong)* to 7 (*Absolutely morally right*). From now on this measure will be referred to as 'moral permissibility' throughout the paper.

Next, participants responded to three items each that measured harm aversion (α = .94; in all studies, Cronbach alphas are calculated across conditions) and perceived suffering capacity of the respective beings (α = .95). A typical harm aversion item was: "How unpleasant would you personally find it to harm the one animal [person]?" with a 9-point response scale ranging from "mildly unpleasant" to "there is nothing else that would be more unpleasant". A typical suffering capacity item was the "How strongly do you think the one animal [person] would suffer if you harm it [him]?" with a 9-point response scale ranging from "mild suffering" to "strongest imaginable suffering".

We then asked participants what the smallest numbers of beings are that would need to be saved to make it permissible to kill the 10 beings (i.e., tipping point). Participants could either respond with a number in a text field or indicate that it is never right. Next, participants completed the Speciesism Scale (Caviola et al., 2019; α = .88), which measures individual differences in the tendency to assign different moral worth based on species-membership. The scale consists of six items, including "Morally, animals always count for less than humans" and "Humans have the right to use animals however they want to". Participants then completed the Empathic Concern Scale (α = .94) and the two subscales of the Oxford Utilitarianism Scale; OUS-IH α = .80 and OUS-IB α = .77). Finally, participants responded to demographic questions.

Results

A one-way ANOVA revealed that, as predicted, moral permissibility differed significantly across species (Figure 2; Table 2), $F(5, 899) = 23.62$, $p < .001$, $\eta^2 = .12$. Participants judged harming humans to save many other humans as more morally wrong than doing so for pandas, dogs, squirrels, chimpanzees and pigs. Similarly, many more participants thought that it was never right to harm ten humans irrespective of the number of saved humans, whereas a much smaller proportion of participants thought so when it concerned animals (Table 2). Figure 2 shows that deontological constraints differed slightly between the animals conditions. However, these differences were only significant between the panda bears and pigs condition and the dogs and pigs condition (Table 3).
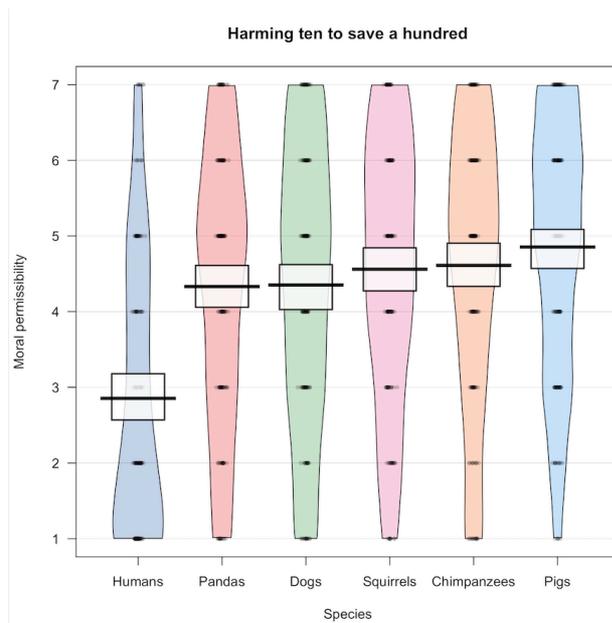
Figure 2. Moral permissibility of harming 10 individuals of a species to saving 100 of the same species, ranging from 1 (*Absolutely morally wrong*), to 4 (*Neither right nor wrong*), to 7 (*Absolutely morally right*). For all figures, black points represent raw data, horizontal bars represent means, rectangles represent confidence intervals, and "beans" represent smoothed densities. (Study 1)

Table 2

Descriptive statistics of moral permissibility and scales (Study 1)

Species	Moral permissibility	Never right	Tipping point	Harm aversion	Suffering capacity
Humans	2.85 (1.84)	65%	100	7.79 (1.53)	7.91 (1.21)
Pandas	4.33 (1.81)	34%	50	7.14 (1.83)	7.33 (1.49)
Dogs	4.35 (1.88)	36%	50	7.38 (1.43)	7.04 (1.77)
Squirrels	4.56 (1.78)	39%	50	6.42 (2.07)	7.01 (1.86)
Chimpanzees	4.35 (1.88)	30%	45	6.70 (2.11)	7.20 (1.71)
Pigs	4.85 (1.69)	25%	50	6.11 (2.08)	6.79 (1.80)

Note. Never right stands for the proportion of participants who thought that it was never right to harm ten individuals of a certain species irrespective of the number of saved individuals of the same species. Tipping point stands for the number of individuals that have to be saved in order to justify harming ten individuals of the same species (median response).

Table 3

Effect sizes (Cohen's d) for moral permissibility comparisons across conditions (Study 1)

	Humans	Pandas	Dogs	Squirrels	Chimpanzees
Pandas	.81***				
Dogs	.81***	.01			
Squirrels	.94***	.13	.11		
Chimpanzees	.96***	.15	.14	.03	
Pigs	1.13***	.30**	.28*	.17	.14

$^†p < .10.$ $^*p < .05.$ $^{**}p < .01.$ $^{***}p < .001.$

Next, we tested the Harm Aversion Mediated Speciesism Model. The relationship between condition and moral permissibility was mediated by harm aversion. The bootstrapped indirect effect was (1.04)(-.31) = -.32, 95% CI [-.43, -.22]. Thus, the indirect effect was statistically significant. We conducted a hierarchical regression to test which factors were driving harm aversion (Table 4). We found that speciesism significantly moderated the effect of condition onto harm aversion. The bivariate correlation between speciesism and harm aversion was -.16 ($p = .05$) in the humans condition and -.48 ($p < .001$) in the animals condition (Table 5). In addition, speciesism itself and perceived suffering capacity significantly predicted harm aversion. Thus, our model was confirmed.

Table 4 shows the regression results when the humans condition was contrasted with all animals conditions collapsed. We also ran separate regressions where we compared the humans condition with each animals condition separately, controlling for perceived suffering capacity. We found that the speciesism moderation effect was .31 (p = .03) for humans vs. pandas, .27 (p = .04) for humans vs. dogs, .34 (p = .005) for humans vs. squirrels, .42 (p < .001) for humans vs. chimpanzees, and .53 (p < .001) for humans vs. pigs. This suggests that the speciesism moderation effect was stronger for those animals that participants grant weaker deontological constraints to.

Finally, we found that the bivariate correlation between speciesism and suffering capacity was stronger in the animals conditions (r = -.40, p < .001) than in the humans condition (r = -.24, p = .003). Speciesism significantly moderated the effect of condition (humans vs. animals) onto suffering capacity (β = .22, p = .003).

Table 4

Hierarchical regression predicting harm aversion (Study 1)

Predictor measures	Step 1	Step 2
R^2	.24	.34
Condition	-.10	-.09
Speciesism	-.49***	-.35***
Speciesism * Condition	.33***	.25***
Suffering		.35***

Note. Displaying standardized coefficients β. Condition was coded as 1 for the humans condition and as 0 for all animals conditions.
†p < .10. *p < .05. **p < .01. ***p < .001.

Table 5

Correlations between measures (Study 1)

	Permissibility	Speciesism	Harm av.	Suffering	EC	OUS-IB	OUS-IH
Speciesism	.35***						
Harm av.	-.36***	-.43***					
Suffering	-.30***	-.37***	.49***				
EC	-.16***	-.35***	.32***	.26***			
OUS-IB	.02	-.27***	.18***	.13***	.37***		
OUS-IH	.42***	.33***	-.19***	-.16***	-.29***	.09**	
Condition	-.33***	-.01	.20***	.18***	.01	.02	-.03

Note. Condition was coded as 1 for the humans condition and as 0 for all animals conditions.
†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

Discussion

The results of Study 1 demonstrate that people consider it significantly more wrong to harm a few humans to save many than to harm a few animals to save many irrespective of the type of animal—whether they are panda bears, dogs, squirrels, chimpanzees or pigs. This suggests that people have weaker deontological constraints for animals than for humans and are therefore more likely to be willing to sacrifice some for the greater good when confronted with animals than with humans. The results supported the Harm Aversion Mediated Speciesism Model.

It was, however, not the case that people considered it completely permissible to harm animals to save many, which is what "Utilitarianism for animals, Kantianism for people" would suggest, nor was the strength of the constraint constant across the scenarios. The results therefore support the Multi-level Weighted Deontology hypothesis on which deontological constraints still limit the permissibility of harm to animals, but in a weaker way than in the human case. Note that Figure 2 reveals that, while there were large individual

differences in responses, only a small minority exhibited *no* deontological constraint at all against harm for animals, which is what Nozick's slogan would predict.

We also found that there are differences in deontological constraints people grant to different types of animal species. For example, people hold stronger deontological constraints against harming dogs (a typical pet) than pigs (a typical food animal). The results of this study are in line with the results of two additional versions of this study we conducted, reported in the Supplementary Materials, in which we only included humans, dogs, pigs, and chimpanzees. The fact that deontological constraints differ (even if only to a small degree) across different animal species is inconsistent not only with Nozick's slogan, but also with the Multi-Level Uniform Deontology view. By contrast, these findings fit well with the Multi-level Weighted Deontology view, according to which deontological constraints get weighted differently depending on how much people value the respective species.

Study 2: Personal vs. impersonal sacrificial dilemmas

In Study 2, we tested whether the effect we found in Study 1—namely that people have weaker deontological constraints for animals than for humans—would replicate when relying on the traditional trolley problem.

In a previous study, which is reported in the Supplementary Materials, we conducted such a study involving pigs and found that participants had weaker deontological constraints for pigs than for humans. In this study, by contrast, we included zebras, an animal that people do not classify as food but rather want to save from extinction (cf. Schubert, Caviola & Faber, 2019). Our hypothesis—pre-registered at https://osf.io/xsbr8—was that participants would find it more permissible to harm one zebra in order to save five zebras than to harm one human to save many humans

According to the dual-process model of moral judgment (Greene, 2014) an automatic and intuitive aversion to harm prevents more deliberate utilitarian cost-benefit analysis. Previous research has shown that in more personal moral dilemmas that involve direct actions (e.g., pushing a person off the bridge) the aversion to harm is more emotional and stronger compared to impersonal moral dilemmas that involve indirect actions (e.g., pushing a button; cf. Greene, 2014). Therefore, one possible conclusion is that harming individuals with lower perceived moral status, such as animals, is less personal and emotionally evocative, resulting in a weaker aversion to harm such individuals.

We, therefore, looked at both personal sacrificial moral dilemmas (Footbridge / Push) and impersonal sacrificial moral dilemmas (Sidetrack / Switch). We intended to test whether deontological constraints for animals are similarly reduced in an impersonal dilemma and in a personal dilemma. Based on the findings of our previous study, we hypothesized that there would be no interaction effect between the type of dilemma and the species-manipulation.

Method

Participants

We recruited 404 US American participants online via MTurk who received $0.50 payment for their participation. Four participants were excluded for either failing to complete the questionnaire or for failing an attention check, leaving a final sample of 400 people (194 females; $M_{age} = 38.36$, $SD = 11.51$). We aimed for a sample size of 400 participants, as our a priori power analysis showed that 351 participants were required to detect a medium effect size of $f = .15$, with an alpha of 0.05, and power of 0.80.

Design, materials and procedure

Our study had a 2 dilemma (impersonal vs. personal) x 2 species (humans vs. zebras) between-subjects factorial design. Participants were randomly allocated to read one of four variants of a moral dilemma modelled after the traditional trolley problem. As a personal dilemma the Footbridge trolley case was used and as an impersonal the Sidetrack trolley case was used. After reading the dilemmas, participants responded to the same scales as in Study 1: harm aversion ($\alpha = .95$), perceived suffering capacity ($\alpha = .93$), Speciesism Scale ($\alpha = .88$), Empathic Concern Scale ($\alpha = .94$), and the Oxford Utilitarianism Scale (OUS-IH $\alpha = .80$; OUS-IB $\alpha = .79$). Finally, participants responded to demographic questions.

Results

A 2x2 between-subjects ANOVA revealed two main effects but no interaction. Participants considered harming one zebra to save five zebras to be more permissible (impersonal: $M = 5.03$, $SD = 1.66$; personal: $M = 3.95$, $SD = 1.93$) than harming one human to save five humans (impersonal: $M = 4.19$, $SD = 1.79$; personal: $M = 2.86$, $SD = 2.02$), $F(1, 396) = 28.08$, $p < .001$, $\eta_p^2 = .06$. Participants also considered harming in an impersonal dilemma (i.e., via flipping a switch) to be more permissible than in a personal dilemma (i.e., via pushing off the footbridge), $F(1, 396) = 41.70$, $p < .001$, $\eta_p^2 = .10$. However, there was no interaction effect between species and dilemma, $F(1, 396) = 0.42$, $p = .52$, $\eta_p^2 = .001$ (Figure 3). This means that the difference in moral permissibility between zebras and humans was roughly the same in both the impersonal, $d = .49$, $p = .008$ (Tukey HSD), and the personal dilemma, $d = .60$, $p < .001$. Similarly, the difference in moral permissibility between the impersonal and personal dilemmas was roughly the same both for zebras, $d = .49$, $p = .008$, and humans, $d = .69$, $p < .001$.
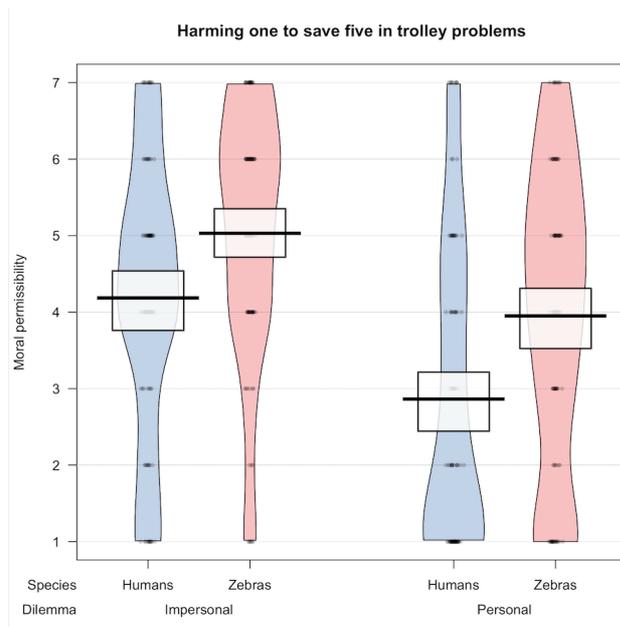
Figure 3. Moral permissibility of harming one individual to saving five individuals of the same species, ranging from 1 (*Absolutely morally wrong*), to 4 (*Neither right nor wrong*), to 7 (*Absolutely morally right*). (Study 2)

Participants were less harm averse for zebras (impersonal: $M = 6.61$, $SD = 1.66$; personal: $M = 6.98$, $SD = 1.77$) than for humans (impersonal: $M = 7.63$, $SD = 1.64$; personal: $M = 7.82$, $SD = 1.59$), $F(1, 396) = 31.23$, $p < .001$, $\eta_p^2 = .07$. Harm aversion was not statistically significantly stronger in personal compared to impersonal dilemmas, $F(1, 396) = 2.75$, $p = .10$, $\eta_p^2 = .007$. There was also no interaction effect, $F(1, 396) = 0.31$, $p = .58$, $\eta_p^2 < .001$. Further, participants perceived zebras (impersonal: $M = 7.66$, $SD = 1.41$; personal: $M = 7.86$, $SD = 1.54$) to suffer less than humans (impersonal: $M = 8.11$, $SD = 1.42$; personal: $M = 8.28$, $SD = 1.48$), $F(1, 396) = 1.62$, $p = .20$, $\eta_p^2 = .02$. Suffering capacity did not differ across the two dilemma types, $F(1, 396) = 2.75$, $p = .10$, $\eta_p^2 = .004$. There was also no interaction effect, $F(1, 396) = 0.009$, $p = .92$, $\eta_p^2 < .001$.

Next, we tested the Harm Aversion Mediated Speciesism Model. The relationship between condition and moral permissibility was mediated by harm aversion. The

bootstrapped indirect effect was (.93)(-.31) = -.29, 95% CI [-.44, -.16]. Thus, the indirect

effect was statistically significant. We conducted a hierarchical regression to test which

factors were driving harm aversion (Table 6). We found that speciesism significantly

moderated the effect of condition onto harm aversion. The bivariate correlation between

speciesism and harm aversion was -.18 (p = .01) in the humans condition and -.43 (p < .001)

in the zebras condition (Table 7). In addition, perceived suffering capacity significantly

predicted harm aversion. Thus, our model was confirmed.

Finally, we found that the bivariate correlation between speciesism and suffering

capacity was stronger in the zebras condition (r = -.28, p < .001) than in the humans

condition (r = -.15, p = .04). Despite this, speciesism did not significantly moderate the effect

of condition onto suffering capacity (β = .18, p = .14).

Table 6

Hierarchical regression predicting harm aversion (Study 2)

Predictor measures	Step 1	Step 2
R^2	.17	.34
Condition	-.02	-.01
Speciesism	-.43***	-.30***
Speciesism * Condition	.34**	.26*
Suffering		.43***

Note. Displaying standardized coefficients β. Condition was coded as 1 for the humans condition and as 0 for the zebra condition.
$^{\dagger}p$ < .10. $^{*}p$ < .05. $^{**}p$ < .01. $^{***}p$ < .001.

Table 7

Correlations between measures (Study 2)

	Permissibility	Speciesism	Harm av.	Suffering	EC	OUS-IB	OUS-IH
Speciesism	.18***						
Harm av.	-.31***	-.30***					
Suffering	-.24***	-.22***	.51***				
EC	-.09†	-.36***	.37***	.26***			
OUS-IB	.16***	-.28***	.19***	.11*	.47***		
OUS-IH	.45***	.25***	-.26***	-.18***	-.15***	.15***	
Condition	-.25***	-.01	.27***	.15***	-.05	-.08	.02

Note. Condition was coded as 1 for the humans condition and as 0 for the zebras condition. †$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

Discussion

Results from Study 2 confirm our main hypothesis that people consider it more permissible to harm one zebra to save five zebras than to harm one human to save five humans. It provides further support that people's deontological constraints for animals are weaker than for humans. The results supported the Harm Aversion Mediated Speciesism Model.

The reduction in deontological constraints for animals compared to humans was similarly strong in both personal and impersonal dilemmas. This is consistent with the findings of a previous study we conducted including pigs, which is reported in the Supplementary Materials. In previous research it has been argued that the difference in responses between personal and impersonal sacrificial dilemmas is the result of a greater emotional reaction in the personal case (Greene, 2014). If so, the lack of interaction could suggest that the mode via which one harms a few to save many (i.e., personal vs. impersonal)

and whether the beings at stake are animals or humans are processed independently and make independent contributions to the overall judgment.

Study 3: Process dissociation

In Study 3, we aimed to more directly test our hypothesis that the deontological aversion to harm animals is weaker than the deontological aversion to harm humans. A limitation of conventional sacrificial moral dilemmas is that they do not allow to clearly distinguish between the deontological aversion to harm the few and the utilitarian desire to help the many. In conventional sacrificial dilemmas these two factors are conflated because the deontological and utilitarian reasons are pitted against each other and result in a single moral judgment. In this study, we systematically teased apart the two psychological tendencies, namely the purely utilitarian inclination (i.e., tendency to maximize positive outcomes factoring out the deontological inclination to reject harm) from the purely deontological inclination (i.e., tendency to reject harm factoring out the utilitarian inclination to maximize the positive outcomes). We did this by relying on the so-called process dissociation technique, originally developed by Jacoby (1991) for the study of memory processes and more recently by Conway and Gawronski (2013) to sacrificial moral dilemmas.

We hypothesized that in particular deontological inclinations would be reduced for animals compared to humans. This is because according to Multi-level Weighted Deontology, deontological constraints get reduced for individuals that people morally value less. In contrast, according to Multi-Level Uniform Deontology, utilitarianism, Cross Species Deontology there should be no difference in the deontological inclination towards humans and animals. In our previous studies, we found evidence for that based on our harm aversion scales. We further hypothesized that utilitarian inclinations would be affected less by the species-membership of the individuals because we assumed that—all else equal and without

any costs or counteracting harm aversion—people would have a similar desire to help many people as to help many animals. We also hypothesized that the more speciesist participants were, the lower their deontological inclinations for animals but not for humans would be.

Method

Participants

We recruited 124 US American participants online via MTurk who received $0.85 payment for their participation. Eleven participants were excluded for failing at least one attention check, leaving a final sample of 113 people (45 females; $M_{age} = 38.19$, $SD = 10.93$). We aimed for a sample size of 110 participants, as our a priori power analysis showed that 110 participants were required to detect a medium effect size of $f = .2$, with an alpha of 0.05, and power of 0.80, and correction among repeated measures of 0.1 (similar to Muda, Niszczota, Bialek, & Conway, 2018).

Design, materials and procedure

The study had two conditions (humans vs. pigs). Participants were presented with ten dilemmas in randomized order involving either pigs or humans. Half of the dilemmas were so-called incongruent dilemmas and half were so-called congruent dilemmas. In the incongruent dilemmas, deontology and utilitarianism prescribe diverging responses, as in the classic footbridge problem (e.g., kill one to save five). In the congruent dilemmas, deontology and utilitarianism prescribe the same responses (e.g., kill five to save one). In order to avoid floor effects in the congruent dilemmas, there were selfish reasons to choose the harmful option (e.g., winning money). The dilemmas can be found in the Supplementary Materials. Participants responded to each dilemma on a 7-point scale, ranging from 1 (*Absolutely morally wrong*) to 7 (*Absolutely morally right*). Responses to the congruent cases (α =.87) and incongruent cases (α = .89) were aggregated from the single dilemmas.

Participants were then presented with questions about the perceived suffering capacity of the respective beings ($\alpha = .97$) and their aversion to harm them ($\alpha = .94$). The questions were similar to the ones asked in the previous studies. Finally, participants responded to the Speciesism Scale ($\alpha = .92$), the Empathic Concern Scale ($\alpha = .95$), the Oxford Utilitarianism Scale (OUS-IH $\alpha = .82$; OUS-IB $\alpha = .81$), and demographic questions.

Results

In the incongruent dilemmas, moral permissibility was higher for pigs ($M = 4.74$, $SD = 1.52$) than for humans ($M = 3.13$, $SD = 1.32$), $t(110) = 6.02$, $p < .001$, $d = 1.13$. This replicates our core finding from the previous studies that people consider it more permissible to harm a few animals to save more animals than they consider it to harm a few humans to save more humans. In the congruent dilemmas, moral permissibility was also higher for pigs ($M = 2.73$, $SD = 1.41$) than for humans ($M = 1.66$, $SD = 0.79$), $t(92) = 4.78$, $p < .001$, $d = .92$.

To compute the utilitarian and deontological parameters, we first transformed the aggregated responses on the 7-point scale into probabilities, indicating the chance that a participant would consider the act immoral ($1 = 100\%$, $4 = 50\%$, $7 = 0\%$). Next, we followed the formula by Conway and Gawronski's (2013) to calculate the two parameters: utilitarian parameter = p(immoral | congruent) – p(immoral | incongruent); deontological parameter = p(immoral | incongruent) / (1 – utilitarian parameter). We standardized the parameters using a z-transformation.

We then conducted a 2(parameter: utilitarian vs. deontological) x 2(species: humans vs. pigs) repeated measures mixed-model ANOVA where the parameter factor was treated within-subjects and the species factor was treated between-subjects. The results revealed a main effect of species-membership, $F(1, 222) = 5.33$, $p = .02$, $\eta_p^2 = .02$, and an interaction effect, $F(1, 222) = 33.86$, $p < .001$, $\eta_p^2 = .13$ (Figure 4). Tukey HSD post-hoc tests revealed

that deontological inclinations were stronger for humans ($M = 0.52$, $SD = 0.65$) than for pigs ($M = -0.48$, $SD = 1.03$), $p < .002$, $d = 1.15$. The difference in utilitarian inclinations between humans ($M = -0.23$, $SD = 0.86$) and pigs ($M = 0.21$, $SD = 1.08$) was not statistically significant, $p = .07$, $d = .44$. However, there was a trend, suggesting that the utilitarian inclination for pigs were descriptively stronger than for humans.
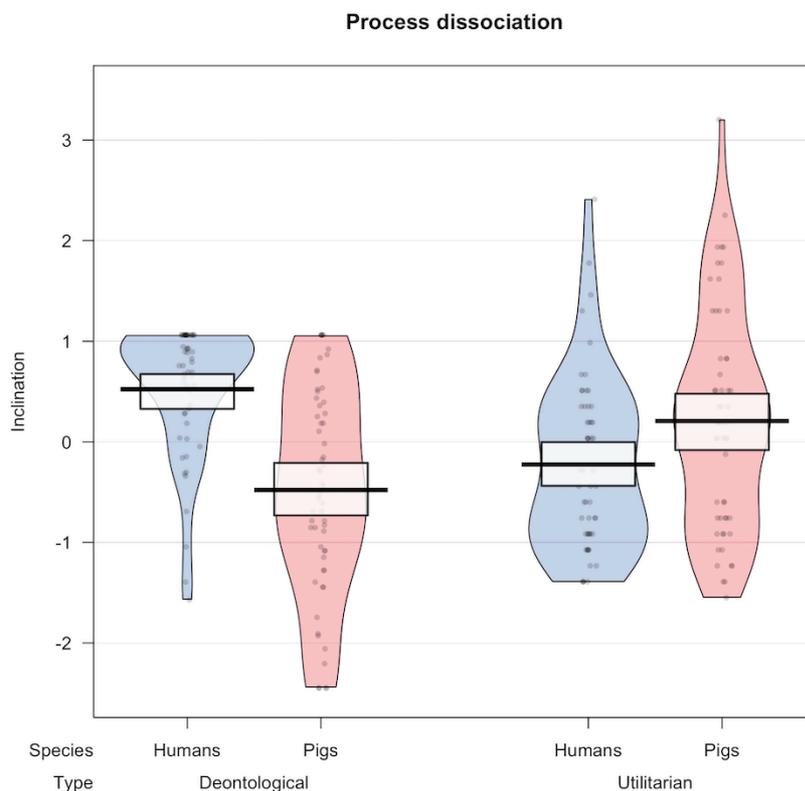
Figure 4. Standardized utilitarianism and deontology parameters for humans and pigs respectively (z-transformed). Participants had stronger deontological inclinations (i.e. harm-rejection) for humans than for pigs, but they had somewhat stronger utilitarian inclinations (i.e. outcome-maximization) for pigs than humans. Note that the zero point is not meaningful due to z-transformation and that a direct comparison between the deontological and utilitarian parameter is not possible. (Study 3)

Next, we looked at correlations across conditions for the utilitarianism and deontology parameter separately (Table 8). In both conditions, harm aversion correlated with the deontological inclination but not with the utilitarian inclination. Perceived suffering capacity correlated with the deontological inclination for pigs but nor for humans. Perceived suffering capacity also did not correlate with the utilitarian inclination in either condition. As hypothesized, the more speciesist participants were, the weaker their deontological inclination for pigs. Speciesism, however, did not correlate with the deontological inclination for humans. Speciesism also did not correlate with the utilitarian inclination in either condition. Across conditions, the utilitarian inclination correlated negatively with the deontological inclination, $r(111) = -.19$, $p = .046$.

Table 8

Correlations with the utilitarian and deontological inclination for each condition separately (Study 3)

Condition	Humans		Pigs	
Inclination	Utilitarian	Deontological	Utilitarian	Deontological
Harm aversion	.04	.33*	.06	.43***
Suffering capacity	-.06	.06	.06	.33*
Speciesism	.03	-.17	-.08	-.65***
Empathic Concern	-.02	.23†	.03	.13
OUS-IB	.01	-.19	.26*	-.01
OUS-IH	.44***	-.52***	.14	-.58***

†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

Next, we tested the Harm Aversion Mediated Speciesism Model. The relationship between condition (humans vs. pigs) and moral permissibility (in the incongruent dilemmas)

was mediated by harm aversion. The bootstrapped indirect effect was $(1.51)(-.20) = -.30$, 95% CI [-.64, -.09]. Thus, the indirect effect was statistically significant. We conducted a hierarchical regression to test which factors were driving harm aversion (Table 9). We found that speciesism significantly moderated the effect of condition onto harm aversion. The bivariate correlation between speciesism and harm aversion was -.13 ($p = .10$) in the humans condition and -.57 ($p < .001$) in the pigs condition (Table 10). In addition, perceived suffering capacity significantly predicted harm aversion. Thus, our model was confirmed.

Finally, we found that the bivariate correlation between speciesism and suffering capacity was -.25 ($p = .06$) in the humans condition and -.55 ($p < .001$) in the pigs condition. However, speciesism did not significantly moderate the effect of condition (humans vs. pigs) onto perceived suffering capacity ($\beta = .32$, $p = .12$).

Table 9

Hierarchical regression predicting harm aversion (Study 3)

Predictor measures	Step 1	Step 2
R^2	.31	.40
Condition	-.02	.08
Speciesism	-.56***	-.38***
Speciesism * Condition	.47*	.35†
Suffering		.33***

Note. Displaying standardized coefficients β. Condition was coded as 1 for the humans condition and as 0 for the pigs condition.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Table 10

Correlations between measures (Study 3)

	z-util	z-deont	Speciesism	Harm av.	Suffering	EC	OUS-IB	OUS-IH
z-deont	-.19†							
Speciesism	-.11	-.36***						
Harm av.	-.03	.50***	-.40***					
Suffering	.02	.18*	-.42***	.45***				
EC	.01	.15	-.18†	.49***	.34**			
OUS-IB	.16†	-.10	-.23*	.20*	.30**	.29**		
OUS-IH	.27**	-.52***	.34***	-.26**	-.16†	-.15*	.24**	
Condition	-.22*	.50***	< .01	.36**	-.03	<. 01	-.07	-.09

Note. Condition was coded as 1 for the humans condition and as 0 for the pigs condition.
†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

Discussion

In Study 3, we systematically teased apart the utilitarian and deontological inclinations, which are conflated in conventional sacrificial dilemma analysis that we have relied on in the other studies. The results show that the deontological inclination is much stronger for humans than for pigs. The utilitarian inclination, by contrast, showed a descriptive trend for being stronger for pigs than for humans. Deontological inclination for pigs was weaker, the lower their perceived suffering capacity and the more speciesist participants were. Perceived suffering capacity also correlated with deontological inclination for humans. Speciesism, however, did not correlate with deontological inclinations for humans. Neither perceived suffering capacity nor speciesism correlated with utilitarian inclinations. The results again supported the Harm Aversion Speciesism Model.

The finding that people had a slightly stronger utilitarian inclination for pigs than humans was somewhat surprising as one would expect people to have a similar desire to save as many humans as possible as to save as many animals as possible. One possibility is that

this result is related to another finding we report below (see Study 5), where we found that people were less likely to agree that one human is less valuable than two humans than they were to agree that one animal is less valuable than two animals. In other words, people may be less willing to maximize helping as many humans as possible than they are willing to maximize helping as many animals as possible (even when there is no counteracting deontological constraint) because they consider it less permissible to quantify and compare the lives of humans than the lives of animals.

In sum, the findings of this study suggest that the fact that people are more willing to harm a few animals to save more animals than to harm a few humans to save more humans is primarily driven by a weaker deontological inclination for animals than humans. And we found that this reduced deontological inclinations for animals is at least partly driven by both perceived suffering capacity and speciesism. This is again in line with our Multi-level Weighted Deontology hypothesis: deontological constraints get weaker, the less people value the individuals; people value animals less than humans, and thus hold weaker deontological constraints for them.

Study 4: Cross-species sacrifices

So far, the individuals that could be saved were of the same species as those that had to be harmed. This means that the perceived moral status of the individuals that have to be harmed and those that can be saved was kept constant. In Study 4, we aim to manipulate the relative value of the costs and benefits. We do this by examining cross-species sacrifices, i.e., cases when the species-membership of the harmed and the saved individuals differs. According to the Multi-level Weighted Deontology model, deontological constraints are weaker the lower the moral status of the individual that has to be harmed. Further, the weaker

the deontological constraints, the stronger the comparative weight of the moral reasons to save the greater number. From this, we draw the following hypotheses:

First, we would expect people to generally consider it more permissible to harm animals than humans irrespective of whether many animals or humans can be saved. This is because in both cases deontological constraints against harming animals are weaker, which makes it more likely for the utilitarian cost-benefit analysis to overrule them. However, since humans have higher moral status than animals, deontological constraints might be even easier to overturn if humans compared to animals can be saved.

Second, since deontological constraints against harming humans are strong, people will in general consider it fairly wrong to harm humans to save either humans or animals. However, to the extent that people overcome their deontological constraints and do engage in cost-benefit analyses, these will yield opposing recommendations. Cost-benefit analysis will recommend harming the human to save many humans but advise against harming a human to save many animals. This is because most people consider one human death to be much worse than five pig deaths. We therefore expect an interaction effect such that the difference in moral permissibility is greater if humans have to be harmed than if animals have to be harmed.

Method

Participants

We recruited 150 US American participants online via MTurk who received $0.90 payment for their participation. Eight participants were excluded for failing at least one of two attention checks, leaving a final sample of 142 people (66 females; $M_{age} = 38.60$, $SD = 11.66$). We anticipated a small to medium effect of $f = .13$ between animal species based on the results of Study 1. With an alpha of 0.05, and power of 0.80, our a priori power analysis

showed that we required 120 participants for a within-subjects study. To account for exclusions, we recruited 150 participants.

Design, materials and procedure

This study had a 2 harm (pigs vs. humans) x 2 save (pigs vs. humans) design. Note that in contrast to the previous studies, this study was conducted completely within-subjects to test whether the effect also holds in this set-up.

We chose an abstract sacrificial dilemma because its briefness was well suited for this within-subjects design and because we expected large effects. Participants were asked to consider the following: "Suppose you are in a situation in which you have to decide whether to kill 10 individuals to save 100 individuals. If you do nothing, the 100 individuals will die. How morally right or wrong is it to... 1) kill 10 humans to save 100 pigs, 2) kill 10 humans to save 100 humans, 3) kill 10 pigs to save 100 pigs, 4) kill 10 pigs to save 100 humans?" Participants responded to these four questions first on a 7-point moral permissibility scale ranging from 1 (*Absolutely morally wrong*) to 7 (*Absolutely morally right*), and on the next page on a binary response measure (*Wrong* vs. *Right*). On the next pages, we included two new questions on moral acceptability and moral requiredness for all four conditions each to test whether they would yield different patterns. Next, we included an question, asking what *minimum* number of individuals would need to be *saved* in order to justify harming ten (similar to the tipping point question in Study 1). Furthermore, we included a question asking what *maximum* number of individuals could be *harmed* in order for it still to be justified to save the hundred. Participants then completed the Speciesism Scale ($\alpha = .84$), the Oxford Utilitarianism Scale (OUS-IB $\alpha = .83$; OUS-IH $\alpha = .71$), and responded to demographic questions.

Results

      A two-way repeated-measures ANOVA revealed two main effects and an interaction. Participants generally considered it more permissible to harm pigs than humans, $F(1, 142) = 461.72$, $p < .001$, $\eta_p^2 = .38$, and generally considered it more permissible to harm an individual if humans could be saved than if pigs could be saved $F(1, 142) = 179.63$, $p < .001$, $\eta_p^2 = .19$. When humans had to be harmed, participants considered this much more permissible if this had to be done in order to save other humans ($M = 4.12$, $SD = 1.70$) than if it had to be done in order to save pigs ($M = 1.83$, $SD = 1.36$), $p < .001$, $d = 1.50$. But, in contrast, when pigs had to be harmed, participants considered this only very slightly more permissible if this had to be done in order to save humans ($M = 5.78$, $SD = 1.54$) than if it had to be done in order to save pigs ($M = 5.04$, $SD = 1.68$), $p < .001$, $d = .47$. When humans could be saved, participants considered it more permissible to do so by harming pigs than by harming humans, $p < .001$, $d = 1.03$. But when pigs could be saved, participants considered it even much more permissible to do so by harming pigs than by harming humans, $p < .001$, $d = 2.10$. In short, there was a significant interaction in moral permissibility based on species harmed and species saved (Figure 5), $F(1, 142) = 46.38$, $p < .001$, $\eta_p^2 = .06$. Results for the moral acceptability and moral requiredness questions were in line with the main dependent variable (see Supplementary Materials).

      The same pattern was found for the binary morally right/wrong question: Only 4.2% considered it right to harm humans to save many pigs, but 51.4% considered it right to harm humans to save many humans. On the other hand, 80.3% considered it right to harm pigs to save many pigs, and 85.9% considered it right to harm pigs to save many humans. 83.8% participants said it would always be wrong to kill humans, no matter the number of saved pigs. 46.5% said it would always be wrong to kill humans, no matter the number of saved humans. In contrast, only 14.8% said it would always be wrong to kill pigs no matter the

number of saved humans. And 18.3% said it would always be wrong to kill pigs no matter the

number of saved pigs.

Harming ten to save a hundred
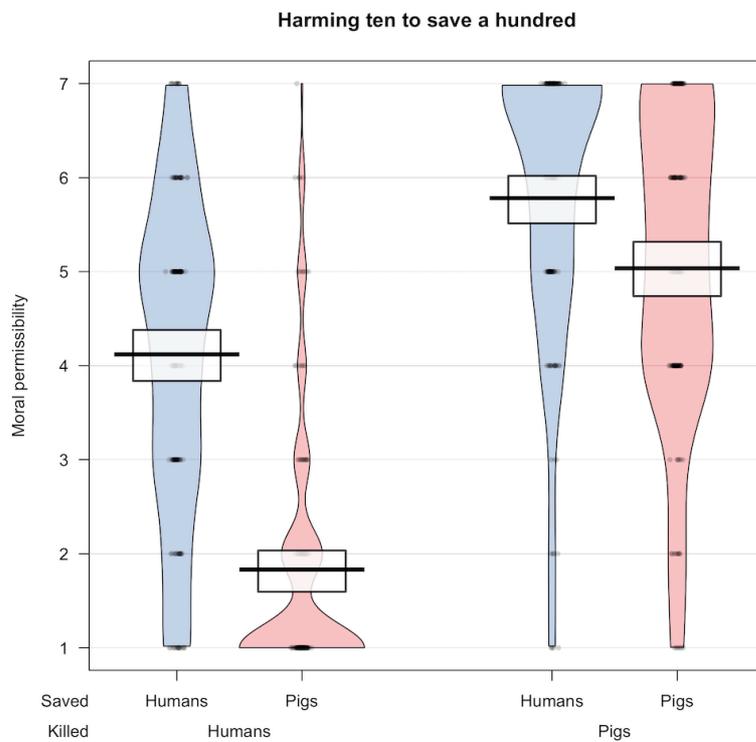
Figure 5. Moral permissibility of harming 10 individuals of a species to saving 100 of either

the same or a different species, ranging from 1 (*Absolutely morally wrong*), to 4 (*Neither*

right nor wrong), to 7 (*Absolutely morally right*). (Study 4)

Discussion

The results of this study confirm our hypothesis. People generally consider it to be

permissible to harm animals irrespective of who will be saved—with only minor differences

in moral permissibility depending on whether humans or animals would be saved. At the

same time, however, people consider it impermissible to harm humans for animals, but less

so if humans can be saved. Harming humans to save more humans—as in the traditional

Footbridge case—remains controversial among participants. Therefore, as we predicted, the

difference in how permissible people consider it to harm humans if either humans or animals

can be saved is much greater than the difference in how moral people consider it to harm

animals if either humans or animals can be saved. Our explanation is that deontological constraints against harming animals are more easily overruled by utilitarian cost-benefit analysis than deontological constraints against harming humans. Further, in the case where humans have to be harmed to save many animals perhaps even cost-benefit analysis advises against the harmful action if people believe that 10 humans have far more value than 100 animals.

A limitation of this study was that it may not have been entirely clear to participants from the short abstract dilemmas whether the individuals that had to be killed would have died anyway. In all other studies we present in this paper, it was made clear that the individuals that have to be killed would not have died otherwise.

Study 5: Between objects and humans

Our previous studies demonstrated that people are more willing to harm a few animals to save many than to harm a few humans to save many. Does this mean that people see animals just as objects that can be sacrificed for any reason? In Study 5, we aimed to compare the extent to which people hold deontological constraints for humans, animals, and objects. In particular, our research question is whether moral permissibility of harming a few animals for many is closer to those of harming a few humans to save many or those of harming (or destroying) a few objects to save many.

Nozick (1974) hypothesized that people view animals as belonging into a moral category between humans and objects. On the one hand, Nozick reasoned, in some ways we treat animals like objects. We consider it permissible to own animals or to harm them for the greater good—something most people now consider to be wrong if it involved humans. On the other hand, since animals can suffer, there are many things we are not allowed to do with animals that we are allowed to do with objects. For example, while it may be permissible to

harm animals for the greater good, it is generally not permissible to unnecessarily harm animals for weaker reasons such as for pure personal pleasure. With objects, however, people appear to mostly consider it permissible to do with them whatever one wants as long as nobody gets hurt or nobody's property rights are violated. A study by Nichols & Mallon, (2006), however, suggested that people apply very weak deontological constraints even to objects. They reported that people apply deontological constraints even to inanimate objects, albeit very weakly: while people did not consider it overall wrong to destroy one object to prevent several objects from being destroyed in a Footbridge-like scenario, they did agree that a moral rule was broken in such a case.

Our hypothesis, which we pre-registered at https://osf.io/aeu3g/, was that people's views about the moral permissibility of harming animals lie somewhere between their views about the moral permissibility of harming (or destroying) humans and objects. This is because, according to the Multi-level Weighted Deontology model, we hypothesize that people have extremely weak deontological constraints for objects, somewhat weak constraints for animals and strong constraints for humans.

The philosopher Martha Nussbaum (1995) has identified a number of dimensions of objectification, such as ownership, instrumentality, denial of autonomy or fungibility (interchangeability). Based on this, we hypothesized these are applied to objects as well as, to some extent, to animals. People are likely to agree that objects are interchangeable, i.e., that one could easily replace one chair or toothbrush by another if they are similar and that they can be owned. The same does not apply to humans since every human is seen as being unique, whereas animals might be in a moral category in between humans and objects. Similarly, people are likely to agree that one object is less valuable than two. In contrast, people are often unwilling to agree that one human is less valuable than two. Again, we hypothesize that animals are in an intermediate moral position. Overall, we expect that moral

permissibility of harming a few to save many might track views on interchangeability and comparability of these objects. This is because harming a few to save many is only a permissible option if one believes that the respective entities are in principle interchangeable and the two options comparable.

Method

Participants

We recruited 603 US American participants online via MTurk who received $0.72 payment for their participation. Forty-six were excluded for failing at least one attention check, leaving a final sample of 557 people (299 females; $M_{age} = 38.39$, $SD = 11.77$). We expected a small to medium effect size. Our analysis showed that 576 participants were required to detect an effect size of $f = .13$ with an alpha of 0.05, power of 0.80, and 3 groups. We aimed to recruit 600 participants to account for any exclusions.

Design, materials and procedure

This study had a between-subjects design with three conditions: humans, animals, and objects. Each participant received two vignettes in random order. The reason we chose two was because we knew from Study 1 that there can be differences between different types of animals. Similarly, we believed that there could be small differences between high and low value objects (cf. Mallon & Nichols, 2006). The two vignettes participants received involved adults and children in the humans condition, dogs and pigs in the animals condition, and paintings and chairs in the objects condition. We were primarily interested in the average effects, which is why we aimed to take the mean of the responses to the two vignettes in each condition as the main dependent variable. The vignettes were based on our vaccine dilemma from the previous studies but were slightly adapted to ensure plausibility and consistency

across the three conditions. Similar to the vaccine dilemma we used in Study 1, in each dilemma the entities that had to be harmed (or destroyed) were of the same type as the entities that could be saved (e.g., destroy 10 paintings to prevent 100 paintings from getting destroyed).

After completing the main task, which involved indicating their moral permissibility ratings about harming a few to prevent similar harm to many in two separate vignettes, participants were presented with short additional questions. First, participants were asked whether they consider it morally acceptable to kill a pig or a human/destroy a chair if one has a strong personal preference to do so. Second, they were asked whether they consider it acceptable to own a chair, pig or human. Third, they were asked whether they think a chair, pig or human can easily be replaced by another one (interchangeability). Fourth, they were asked whether they believe one chair, pig, or human is less valuable than two (comparability). Next, participants were explicitly asked to indicate on a slider what moral value they attribute to animals relative to objects (-50) and humans (+50). Finally, participants completed the Speciesism Scale (α =.86), the Oxford Utilitarianism Scale (OUS-IB α = .79; OUS-IH α = .76), and responded to demographic questions.

Results

Within each condition, moral permissibility of the two vignettes correlated strongly with each other (humans: $r(187) = 0.95$, $p < .001$; animals: $r(189) = .87$, $p < .001$; objects: $r(175) = .74$, $p < .001$). Therefore, the responses of the two vignettes in each condition were averaged to form a single score. Further analysis revealed that if the two vignettes were analyzed separately, the overall pattern of the results remained the same. The following analyses are based on the aggregated scores only.

A one-way ANOVA revealed that there were significant differences across the three conditions in moral permissibility, $F(2, 554) = 105.67$, $p < .001$, $\eta_p^2 = .28$ (Figure 6; Table 11). And as in our previous studies, participants considered it more permissible to harm a few animals to save many animals ($M = 4.70$, $SD = 1.66$) than to harm a few humans to save ($M = 3.28$, $SD = 2.00$), $p < .001$, $d = .78$. Participants considered it more permissible to destroy a few objects in order to prevent more objects from being destroyed ($M = 5.83$, $SD = 1.28$) than to harm a few animals to save more animals, $p < .001$, $d = .75$. Accordingly, the difference in moral permissibility between the objects and humans condition was very large, $p < .001$, $d = 1.51$.
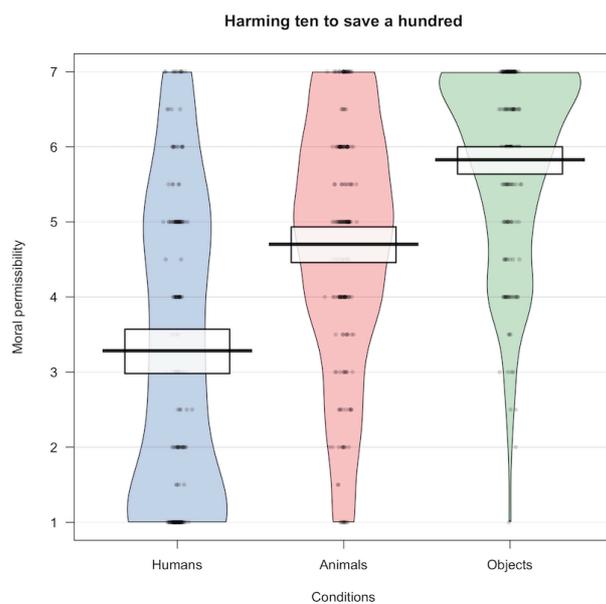

Figure 6. Moral permissibility of harming (or destroying) 10 humans, animals, or objects to save 100 of the same type, ranging from 1 (*Absolutely morally wrong*), to 4 (*Neither right nor wrong*), to 7 (*Absolutely morally right*). (Study 5)

Table 11

Descriptive statistics of moral permissibility and scales (Study 5)

	Humans	Animals	Objects
Harming a few to save many	3.28 (2.00)	4.70 (1.66)	5.83 (1.28)
Harming for personal preference	1.56 (1.23)	3.16 (1.72)	4.95 (1.53)
Owning	1.43 (1.07)	5.76 (1.24)	6.47 (0.9)
Replacing with another (interchangeability)	1.75 (1.19)	4.46 (1.7)	5.98 (1.03)
One less valuable than two (comparability)	2.35 (1.66)	4.48 (1.83)	5.36 (1.60)

Note. M (SD) ratings on a scale from 1 (*Absolutely morally wrong*) to 7 (*Absolutely morally right*) for the first row and from 1 (*Strongly disagree*) to 7 (*Strongly agree*) for the remaining rows.

Next, we looked at the other dependent measures. For all four questions we found the same pattern, with animals being in a category between humans and objects (Table 11). First, participants thought it was permissible to destroy a chair if one has a strong preference to do so, but it was somewhat wrong to harm a pig if one has a strong preference to do so, and very wrong to harm a human if one has a strong preference to do so. Second, participants thought it was permissible to own a chair, as well as a pig—even though somewhat less—but they thought it was wrong to own a human. Third, they thought that a chair could easily be replaced by another one, whereas a human could not be replaced by another human. For pigs, they thought it was somewhat possible to replace a pig with another one. Fourth, participants agreed that one chair was less valuable than two, slightly agreed that one pig was less valuable than two, but strongly disagreed that one human was less valuable than two. When asked explicitly what the moral value of animals relative to humans and objects were, participants stated that they consider animals to be closer to humans with a mean of 23.68 (*SD* = 23.54) on a scale from -50 (objects) to +50 (humans), which was significantly above the mid-point 0, $t(556) = 23.74$, $p < .001$, $d = 1.01$.

Discussion

This study showed that people place animals into a moral category between objects and humans. People have very strong deontological constraints for humans, relatively weak ones for animals, and extremely weak ones, at best, for objects. These findings support the Multi-level Weighted Deontology model. They suggest that there is a hierarchy of moral levels, and that deontological constraints do not disappear as we move down the levels but just get weaker. It is noteworthy that while the modal response in the objects condition was 7 (i.e., no deontological constraints at all), a considerate number of participants (155 out of 177) chose values below 7. And while this could be partly be explained by noise, it is also possible that some people still hold weak deontological constraints for objects.

Similar patterns were found on views about harming for idiosyncratic reasons, ownership, interchangeability, and comparability. Participants considered it wrong to own humans, but morally acceptable to own animals and objects. They considered it very wrong to harm humans to fulfil a personal preference, fairly wrong to harm animals to fulfil a personal preference but morally acceptable to harm objects to fulfil a personal preference. Furthermore, we found that deontological constraints against harming for the greater good may be related with—or even underpinned by—views on interchangeability and comparability of the respective entities. Participants disagreed that humans can easily be exchanged by other humans or that one human life is less valuable than two. At the same time, they had strikingly different views about objects: they believed that objects can easily be replaced by other objects and that one object is clearly less valuable than two objects. Again, animals were placed in an intermediate position with regard to these two aspects. Thus, our results suggest that only if one believes that an entity can be compared with and replaced by other similar entities, is it seen as sensible to harm one such entity to save many.

Study 6: Time investment and donations

The previous studies focused on judgments in hypothetical scenarios and were presented to online samples on MTurk. In Study 6, we aimed to replicate the effect with an offline student sample with two goals: first, to show that the effect shows in the real-world context of medical experimentation; second, to show that the hypothetical judgments we studied so far translate into concrete behavior. In the real world, harming animals to benefit others is common in the context of medical experimentation. The general societal consensus is that medical experimentation on animals, even if harmful, is justified whereas on humans it is not—in particular if no consent is given by the humans.

In this study, we asked students on the University of Oxford campus about their opinion on the ethical permissibility of medical experiments on pigs or human infants. In addition to measuring their opinions in the form of judgments, we also aimed to measure behavior. For that, we gave them the possibility to invest their personal time to help improve the campaign of an activist organization that was fighting against the proposed medical research program on pigs or infants. Furthermore, participants had the option of donating personal money to support the activist organization if they wished to do so. Our hypothesis was that we would replicate the effect found in the previous studies for the judgment as well as behavioral measures.

Method

Participants

We recruited 208 students on the campus of the University of Oxford who received £3 payment (in line with UK minimum wage) for their participation. Eight were excluded because they did not complete an attention check or did not finish the study, leaving a final sample of 200 people (118 females; $M_{age} = 24.31$, $SD = 7.68$). Power analysis showed that 197 participants were required to detect an effect size of $f = .2$, which we determined in an

online pilot study, with an alpha of 0.05, and power of 0.80. We aimed to recruit 200

participants to account for any exclusions.

Design, materials, and procedure

The study had two between-subjects conditions. Participants first read a text about a

planned medical experiment at the university's medical science lab. The research was

described as involving experimentation on either 50 young pigs or human infants, depending

on the condition, in order to develop a medicine that would help thousands of sick pigs or

human infants. The text stated that while the research could be painful for the test subjects, it

would not have any long-term negative side effects. Next, participants were informed that

due to ethical concerns of this planned medical research program the advocacy group

Unethical Research Watch plans to fight against the implementation of the research. In order

to do so, the advocacy group asked us to conduct a survey to assess the opinions of the

general public, whose results will inform their campaign. As part of this alleged survey,

participants were asked to indicate whether they consider the planned research program to be

ethically justified or not, and whether they would like it to stop. On the next page,

participants were told that if they shared the advocacy group's view and wanted to support

their campaign, they could do so by completing a short writing task. In particular, participants

were told that the advocacy group is looking for the best arguments to support their view,

which they could write down on the paper if they wished to support the campaign, and that

we would forward their responses to the advocacy group in order for them to improve their

campaign. Finally, participants were also given the opportunity to donate any amount

between 0 and 300 pennies of their £3 payment to support the advocacy group.

Participants also completed the Speciesism Scale ($\alpha = .73$) and Oxford Utilitarianism

Scale (OUS-IH $\alpha = .80$; OUS-IB $\alpha = .71$), as well as demographic questions. At the end of

the study, participants were debriefed about the fact that both the planned medical research

program as well as the advocacy group were made up. Participants kept the money that they

intended to donate.

Results

Participants considered the proposed medical research program significantly more

ethically justified if it involved pigs (M = 5.39; SD = 1.34) compared to human infants (M =

4.34; SD = 1.77; Figure 7), $t(183)$ = -4.71, p < .001, d = .67. Similarly, they were

significantly more likely to say that they wanted the program to stop if it involved infants (M

= 3.49; SD = 1.88) compared to pigs (M = 2.69; SD = 1.50), $t(189)$ = 3.35, p < .001, d = .48.

Next, we looked at the behavioral measures. Significantly more participants were willing to

invest time to help to support the campaign of the activist organization fighting against the

medical research program if the research involved human infants (38%) than pigs (19%),

$\chi^2(1)$ = 7.71, p = .005. Finally, participants also donated significantly more money (pennies)

to support the activist group if the research involved human infants (M = 51.52; SD = 107.96)

than pigs (M = 21.15; SD = 70.27), $t(169)$ = 2.34, p = .02, d = .33. Since donations were not

normally distributed, we conducted a Wilcoxon rank test, which supported the finding, W =

5348.5, p = .03.
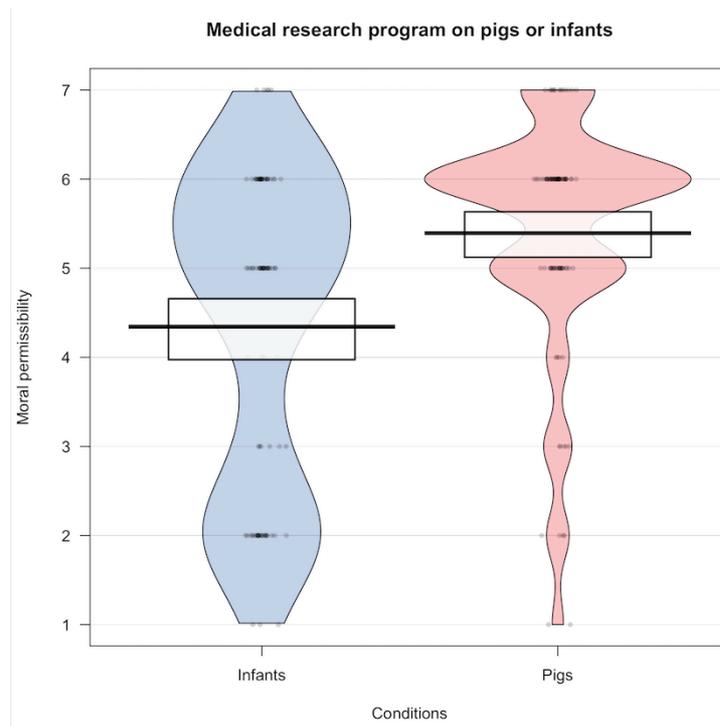
Figure 7. Moral permissibility of the medical research program that harms pigs to save many more pigs or human infants to save many more human infants, ranging from 1 (*Completely unjustified*), to 4 (*Neither justified nor unjustified*), to 7 (*Completely justified*). (Study 6)

Discussion

This study replicated the effect we found in the previous studies in an offline student sample, namely that people have weaker deontological constraints for animals than for humans. This shows that the effect exists in the real-world context of medical experimentation and, crucially, that it translates into behavior. People were more willing to invest both their time and personal money to support an activist campaign against medical research on human infants than on pigs. Nevertheless, there was some degree of opposition to research on animals even when this would help thousands of other animals. This is in line with the Multi-level Weighted Deontology model that suggests that deontological constraints get weaker, but do not disappear, the lower the perceived moral status of the respective individual.

Figure 7 suggests that responses in the infants condition were bi-modal. While one group of participants considered the medical research program mostly justifiable, another group considered it mostly unjustifiable. In the pigs condition, in contrast, all participants considered it justifiable. Since we did not find the same bi-modal pattern in any of our previous studies, it is likely that the pattern is specific to this study. It may be that some participants considered this particular medical research project less harmful than others. One possible reason is that the ratio of harmed and help beings was bigger (harm fifty to help thousands) in this study than in previous studies (e.g., harm ten to help one hundred).

Study 7: Social connectedness

While we found support for the Harm Aversion Mediated Speciesism Model in our previous studies, it might be objected that this model leaves out a potentially important factor. That is the possibility that people are more willing to harm animals because they feel less socially connected to them (Petrinovich et al., 1993). Since people generally feel less socially connected to animals than to other humans, this possibility offers an alternative explanation of the effect we observed in the previous studies. In Study 7, we tested this hypothesis.

To test that whether social connectedness could explain our effect, we employed a 2 species (humans vs. dogs) x 2 social connectedness (close vs. distant) between-subjects study design. Our hypothesis was that participants would continue to hold stronger deontological constraints against harming dogs than humans even when social connectedness is held constant.

Method

Participants

We recruited 302 US American participants online via MTurk who received $0.40 payment for their participation. One participant was excluded for failing an attention check, leaving a final sample of 301 people (143 females; M_{age} = 41.69, SD = 12.85). We aimed for a sample size of 300 participants, as our a priori power analysis showed that 301 participants were required to detect a medium effect size of f = .18, with an alpha of 0.05, and power of 0.80.

Design, materials, and procedure

Participants in the close conditions were first told that they should imagine that all six people (or dogs) described in the scenario were their own brothers (or dogs) that they care very much about. In the distant conditions, they were told to imagine that all six people (or dogs) came from a country far away, that they have never met them before and will never meet them again in the future. Then participants were presented with a moral dilemma similar to the vaccine dilemma used in Study 1 in which they were asked whether they would actively infect one unaffected individual, and by doing so killing it, in order to save five others of the same type. Like in Study 1, participants responded on a 7-point scale ranging from 1 (*Absolutely morally wrong*) to 7 (*Absolutely morally right*).

On the next pages, participants were presented with follow-up questions. First, we measured perceived social connectedness (α = .99), which served as a manipulation check using three items, such as "How socially connected do you feel with the brothers / strangers / dogs (described in the previous scenario)?" on 9-point response scale from *not at all socially connected* to *extremely socially connected*. Next, we measured harm aversion (α = .95) and perceived suffering capacity (α = .92) adjusted to the respective experimental condition with

similar measures as in Studies 1 and 2. Finally, participants responded to the Speciesism

Scale ($\alpha = .88$), the Empathic Concern Scale ($\alpha = .94$), and demographic questions.

Results

      A two-way ANOVA revealed a main effect for the species factor but neither main

effect for the social connectedness nor an interaction effect (Table 12). In line with our

previous studies, participants considered it significantly more permissible to harm one dog to

save five dogs than to harm one human to save many humans, $F(1, 297) = 31.10$, $p < .001$,

$\eta_p^2 = .09$ (Figure 8). The social connectedness factor made no difference, $F(1, 297) = 0.02$, p

$= .90$, $\eta_p^2 < .001$. Participants considered it roughly equally permissible to harm their own

brother to save five other brothers as to harm a stranger to save five strangers, $p = 0.79$, $d =$

.14. Similarly, participants considered it roughly equally permissible to harm their own dog

to save five other own dogs as to harm a dog they don't know to save five other dogs they

don't know, $p = .69$, $d = .19$. As such, there was no significant interaction effect between

species and social connectedness, $F(1, 297) = 2.02$, $p = .16$, $\eta_p^2 = .006$.

      Social connectedness ratings were significantly lower (i.e., less close) in the distant

compared to in the close conditions, $F(1, 297) = 241.73$, $p < .001$, $\eta_p^2 = 0.45$, which serves as

a manipulation check. They were also lower for humans than dogs, $F(1, 297) = 11.68$, $p <$

.001, $\eta_p^2 = 0.04$. And the difference in perceived social connectedness between brothers and

strangers was greater than the difference in perceived social connectedness between one's

own dogs and dogs one doesn't know, $F(1, 297) = 7.64$, $p = .006$, $\eta_p^2 = .03$.

      Harm aversion ratings were stronger in the close compared to the distant conditions,

$F(1, 297) = 15.82$, $p < .001$, $\eta_p^2 = 0.05$. Harm aversion for dogs was weaker than harm

aversion for humans, $F(1, 297) = 4.89$, $p = .03$, $\eta_p^2 = 0.02$. There was no significant

interaction effect for harm aversion, $F(1, 297) = 2.72$, $p = .10$, $\eta_p^2 = .009$. Suffering capacity

ratings were lower for dogs than for humans, $F(1, 297) = 12.70$, $p < .001$, $\eta_p^2 = 0.04$. There

was neither an effect of social connectedness on suffering capacity, $F(1, 297) = 1.74$, $p = .19$,

$\eta_p^2 = .005$, nor was there a significant interaction effect for suffering capacity, $F(1, 297) =$

0.50, $p = .48$, $\eta_p^2 = .002$.

Table 12

Descriptive statistics of moral permissibility and scales (Study 7)

Species	Humans		Dogs	
Social connectedness	Close	Distant	Close	Distant
Moral permissibility	3.32 (1.97)	3.04 (1.95)	4.22 (1.78)	4.55 (1.81)
Social connectedness	7.28 (2.15)	2.67 (2.11)	7.45 (1.87)	4.23 (2.53)
Harm aversion	7.71 (1.88)	7.22 (1.87)	7.60 (1.55)	6.41 (1.98)
Suffering capacity	7.4 (1.37)	7.29 (1.60)	6.87 (1.75)	6.50 (1.71)

Note. M (SD) Moral permissibility on a scale from 1 (wrong) to 7 (right), social
connectedness on a scale from 1 (distant) to 9 (close), harm aversion on a scale from 1
(mild) to 9 (maximum), suffering capacity on a scale from 1 (mild) to 9 (maximum).
Displayed results for social connectedness, harm aversion and suffering capacity are
aggregated responses of three items each.
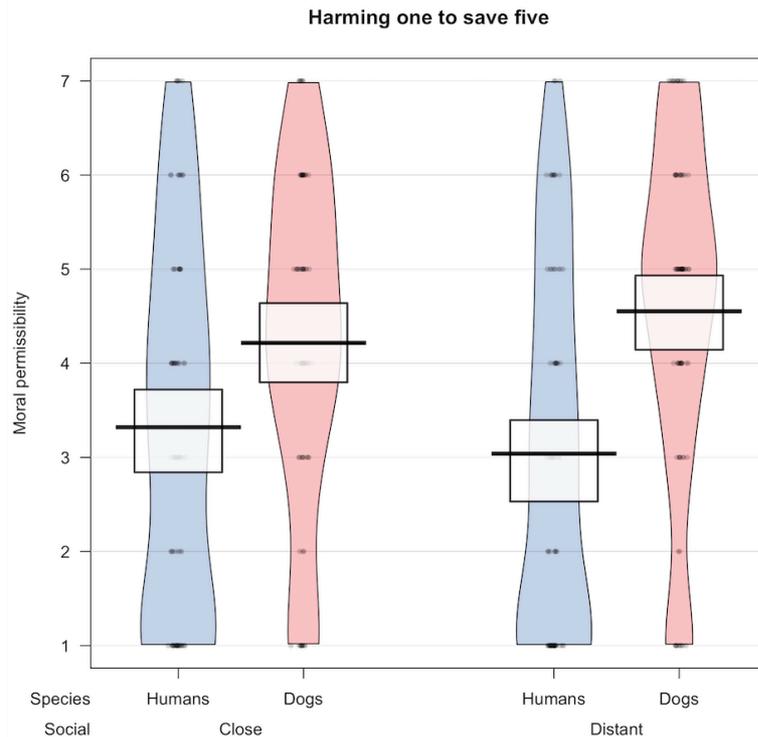
Figure 8. Moral permissibility of harming one to save five individuals, ranging from 1 (*Absolutely morally wrong*), over 4 (*Neither right nor wrong*), to 7 (*Absolutely morally right*). Close humans were described as one's own brothers, close dogs as one's own dogs, distant humans as strangers, distant dogs as dogs one does not know. (Study 7)

Next, we tested the Harm Aversion Mediated Speciesism Model. The relationship between condition (humans vs. dogs) and moral permissibility was mediated by harm aversion. The bootstrapped indirect effect was (.47)(-.19) = -.09, 95% CI [-.21, -.02]. Thus, the indirect effect was statistically significant. We conducted a hierarchical regression to test which factors were driving harm aversion (Table 13). We found that speciesism did not significantly moderate the effect of condition onto harm aversion. The bivariate correlation between speciesism and harm aversion was -.13 ($p = .10$) in the humans condition and -.25 ($p = .001$) in the dogs condition (Table 14). However, perceived suffering capacity significantly predicted harm aversion. Thus, our model was only partly supported.

Finally, we found that the bivariate correlation between speciesism and suffering capacity was stronger in the dogs condition ($r = -.32$, $p < .001$) than in the humans condition ($r = -.14$, $p = .09$). Speciesism significantly moderated the effect of condition onto suffering capacity ($\beta = .34$, $p = .03$).

Table 13

Hierarchical regression predicting harm aversion (Study 7)

Predictor measures	Step 1	Step 2
R^2	.10	.36
Condition (Species)	-.06	-.02
Social connectedness	.23***	.18***
Speciesism	-.29***	-.10
Speciesism * Condition	.24	.05
Suffering		.54***

Note. Displaying standardized coefficients β. Condition was coded as 1 for humans and as 0 for dogs. Social connectedness was coded as 1 for close and as 0 for distant.
†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

Table 14

Correlations between measures (Study 7)

	Permissibility	Speciesism	Harm av.	Suffering	EC
Speciesism	.09				
Harm av.	-.21***	-.18***			
Suffering	-.36***	-.21***	.58***		
EC	-.07	-.20***	.25***	.19***	
Condition	-.31***	.08	.12*	.20***	-.01

Note. Condition was coded as 1 for humans and as 0 for dogs.
†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

Discussion

The results of this study suggest that social connectedness cannot explain why people hold weaker deontological constraints for animals than humans. Instead, other factors associated with species-membership or species-membership itself must explain it. This suggests that the Harm Aversion Mediated Speciesism Model does not need to be extended by including a social connectedness factor.

Despite this, the Harm Aversion Mediated Speciesism Model was only partly supported by our results. Similar to Studies 1, 2, and 3, we found that harm aversion mediated the effect of condition to moral permissibility. And we again found that perceived suffering capacity drove harm aversion. However, in contrast to the previous studies, we found no significant moderation of speciesism on the relationship between condition and harm aversion. One possible explanation is that we used dogs in the current study. Recent research has shown that even speciesists can have a strong emotional attachment to dog, a paradigmatic pet animal (Caviola & Capraro, 2020)—especially if it is their own dog. Thus, the effect of speciesism on the aversion to harming dogs might be less pronounced than its effect on the aversion to harming other animals, such as pigs. Indeed, in Study 1 we found that the moderation effect was weaker for dogs than for pigs. It thus may well be that the effect would have been significant had we used pigs instead of dogs. However, since the correlation between speciesism and aversion to harm dogs was stronger ($r = -.25$) than between speciesism and aversion to harm humans ($r = -.13$), it is also possible that a moderation effect could be detected with a higher sample size. Either way, the results of this study demonstrate the robust relationship between species-membership, perceived suffering

capacity, and harm aversion, which determines moral permissibility of harming a few for the greater good. Finally, despite the fact that speciesism did not significantly moderate the effect of condition onto harm aversion, it did significantly moderate the effect of condition onto perceived suffering capacity. That is, the more speciesist participants were the weaker they perceived the suffering capacity of dogs to be, which could in turn lead to reduced aversion to harm dogs.

Kurzban, DeScioli and Fein (2012) found that people are more willing to harm their brother to save five brothers than they are to harm a stranger to save five strangers. In our study, we did not replicate these findings. Despite the fact that participants rated their social connection to the strangers as much more distant than to their brothers (as per our manipulation check), participants considered it equally permissible to harm one brother to save five brothers than to harm one stranger to save five strangers. One might even wonder whether Multi-level Weighted Deontology would not predict that people should have weaker deontological constraints for strangers than for brothers, since people arguably care more about their brothers. However, it is unlikely that people believe that their brothers have a higher moral status than other human beings in an absolute sense. Instead, they rather believe that they have a special obligation to their brothers than to strangers. It is possible to believe that one has special (agent-relative) obligations to certain people while still believing that all people have the same absolute moral status, and that they therefore possess the same rights. By contrast, as we have shown in other work, people—at least to some extent—believe that humans have a higher moral status than animals in an absolute sense (Caviola et al., 2020).

Study 8: Cognitive capacity and 'marginal cases'

Humans typically possess a more advanced cognitive capacity than animals. In philosophical discussion, the advanced cognitive capacity of humans—enabling language

use, complex thought, and explicit reasoning and deliberation—is often seen as the basis for rational agency or 'personhood' and thus as a central source of moral status. On Kant's view, which we discussed earlier (Kant, 1785), rational agency is the only basis for moral status. But even those who reject Kant's extreme view often accept that rational agency endows a being with superior moral status and is the only basis for moral rights, including deontological protections against being harmed in order to promote a greater good (cf. Gruen, 2017). It is thus possible that people grant humans stronger deontological constraints because they possess such an advanced cognitive capacity that animals lack. In Study 8, we tested this hypothesis.

One way to test this hypothesis is to look at so-called 'marginal cases', as they are used in philosophical discussions (Singer, 1993). These are cases in which humans, such as infants or cognitively severely impaired people, have a cognitive capacity that is similar to, or even lower than, that of some animals. If it can be demonstrated that people grant stronger deontological constraints to humans than animals, even when these humans have lower cognitive capacity than the respective animals, this would suggest that cognitive capacity cannot be a sufficient explanation of the effect in question. In our study, we made use of such marginal cases in our experimental paradigm. Those who appeal to this example tend to interpret the prioritization of marginal cases over animals as showing that people are speciesist (Singer, 1975).

Method

Participants

We recruited 238 US American participants online via MTurk who received $0.45 payment for their participation. Five participants were excluded for failing an attention check,

leaving a final sample of 233 people (94 females; $M_{age} = 37.69$, $SD = 10.99$). We aimed for a

sample size of at least 210 participants.

Design, materials and procedure

Participants were again presented with a modified version of the vaccine dilemma

from the previous study. There were two conditions: patients and chimpanzees. In the

chimpanzees condition, participants were asked to imagine a scenario where there are two

shelters that host chimpanzees. The chimpanzees were described as being relatively

intelligent, in comparison to other animals: as having ways of communicating and of forming

social relationships, and as being able, in a limited way, to plan for the future and make

autonomous, informed decisions. It was also stated that some researchers believe that

chimpanzees are self-aware. In the patients conditions, participants were asked to imagine

that there are two clinics that host severely cognitively impaired humans. These humans were

described as having very limited intelligence, even lower than those of the chimpanzees in

the other condition. Finally, in both conditions it was made clear that the chimpanzees (or

patients, respectively) are capable of experiencing physical and emotional pain. Next, as in

our prior vaccine dilemmas, participants were asked whether they consider it morally right or

wrong to actively kill 10 unaffected chimpanzees (patients) to save 100 other chimpanzees

(patients) on a 7-point scale.

Participants were then presented with questions about the perceived cognitive

capacity (intelligence, rationality, and capability of planning into the future; $\alpha = .93$) of the

respective beings, their perceived suffering capacity ($\alpha = .96$) and aversion to harm these

beings ($\alpha = .91$). Finally, participants responded to the Speciesism Scale ($\alpha = .90$), the

Empathic Concern Scale ($\alpha = .89$), and demographic questions.

Results

   Chimpanzees were perceived to have a higher cognitive capacity ($M = 6.01$; $SD = 1.50$) than severely cognitively impaired humans ($M = 3.74$, $SD = 2.38$), $t(196) = 8.76$, $p < .001$, $d = 1.45$, which served as a manipulation check. Participants considered it more permissible to harm ten chimpanzees in order to save 100 chimpanzees ($M = 4.43$, $SD = 1.79$) than to harm 10 severely cognitively impaired humans to save 100 severely cognitively impaired humans ($M = 3.47$, $SD = 2.02$; Figure 9), $t(228) = 3.85$, $p < .001$, $d = .50$. Harm aversion for chimpanzees ($M = 6.62$; $SD = 1.78$) was significantly lower than for patients ($M = 7.10$, $SD = 1.85$), $t(231) = -2.01$, $p = .045$, $d = .26$. Perceived suffering capacity for chimpanzees ($M = 5.20$, $SD = 2.51$) was not significantly different than for patients ($M = 5.11$, $SD = 2.57$), $t(231) = 0.27$, $p = .79$, $d = .04$.
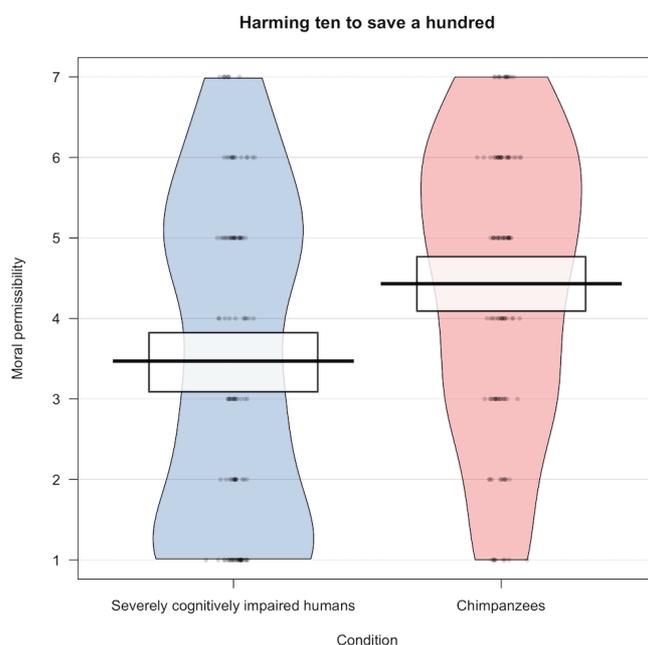

Figure 9. Moral permissibility of harming ten chimpanzees to save a hundred chimpanzees or of harming ten severely cognitively impaired humans (with a lower cognitive capacity than the chimpanzees) to save a hundred severely cognitively impaired humans, ranging from 1

(*Absolutely morally wrong*), over 4 (*Neither right nor wrong*), to 7 (*Absolutely morally right*). (Study 8)

Next, we tested the Harm Aversion Mediated Speciesism Model. The relationship between condition (patients vs. chimpanzees) and moral permissibility was mediated by harm aversion. The bootstrapped indirect effect was (.48)(-.31) = -.15, 95% CI [-.33, -.02]. Thus, the indirect effect was statistically significant. We conducted a hierarchical regression to test which factors were driving harm aversion (Table 15). We found that speciesism significantly moderated the effect of condition onto harm aversion. The bivariate correlation between speciesism and harm aversion was -.20 (p = .03) in the cognitively impaired humans condition and -.43 (p < .001) in the chimpanzee condition (Table 16). In addition, both perceived suffering capacity (positively) and perceived cognitive capacity (negatively) predicted harm aversion. Thus, the Harm Aversion Mediated Speciesism Model was confirmed.

Finally, we found that the bivariate correlation between speciesism and perceived cognitive capacity was -.14 (p < .12) in the chimpanzees condition and .41 (p < .001) in the humans condition. Speciesism significantly moderated the effect of condition onto perceived cognitive capacity (β = .68, p < .001). And we found that the bivariate correlation between speciesism and suffering capacity was -.09 (p = .33) in the chimpanzees condition and .16 (p = .08) in the humans condition. Speciesism did not significantly moderate the effect of condition onto suffering capacity (β = .33, p = .06).

Table 15

Hierarchical regression predicting harm aversion (Study 8)

Predictor measures	Step 1	Step 2

R^2	.11	.16
Condition	-.15	-.26
Speciesism	-.43***	-.42***
Speciesism * Condition	.30†	.34*
Suffering		.20*
Cognitive capacity		-.16*

Note. Displaying standardized coefficients *β*. Condition was coded as 1 for cognitively impaired humans and as 0 for chimpanzees.
†*p* < .10. *p* < .05. **p* < .01. ***p* < .001.

Table 16

Correlations between measures (Study 8)

	Permissibility	Speciesism	Harm av.	Suffering	Cogn. capacity	EC
Speciesism	.31***					
Harm av.	-.32***	-.32***				
Suffering	-.04	.04	.18**			
Cogn. Capacity	.43***	.20***	-.17**	.17**		
EC	-.25***	-.43***	.26***	-.02	-.18**	
Condition	-.25***	-.07	.13*	-.02	-.50***	.11†

Note. Condition was coded as 1 for cognitively impaired humans and as 0 for chimpanzees.
†*p* < .10. *p* < .05. **p* < .01. ***p* < .001.

Discussion

This study showed that people hold stronger deontological constraints against harming humans than against harming animals even in cases where they perceive the humans

to have lower cognitive capacity than the animals. In line with the Harm Aversion Mediated Speciesism Model, participants were more averse to harm the humans than the chimpanzees. This was the case despite the fact that people perceived them both to be similarly capable of suffering. We also found, in line with the model's predictions, that speciesism moderated the effect of condition onto harm aversion. That is, the more speciesist participants were, the less averse they were to harm the chimpanzees in particular. In line with the model, perceived suffering capacity and perceived cognitive capacity were also significant predictors of harm aversion. Overall, these results confirm our model. They show that the reduced harm aversion for animals is multiply determined and cannot be fully explained by perceived cognitive capacity alone.

Study 9: Cognitive capacity in a hypothetical species

In Study 8 we found that people hold stronger deontological constraints against harming humans than animals even in cases where they perceive the humans to have a lower cognitive capacity than the animals. This suggests that other factors must explain the effect. However, it does not rule out that cognitive capacity still partly contributes to the effect. In Study 9, we examined whether cognitive capacity partly drives deontological constraints against harm if everything else—including species—is held constant.

To test this, we presented participants with scenarios involving a hypothetical extraterrestrial species who either had advanced or basic cognitive capacity and who either had a low or a high degree of suffering capacity. Our study was inspired by a study conducted by Sytsma and Machery (2012) who found that participants considered it more wrong to harm cognitively advanced, compared to cognitively basic, beings to benefit humans.

Our hypothesis, which we pre-registered at https://osf.io/24vwr/, was that participants consider harming individuals with an advanced cognitive capacity (just like humans) to be wrong but individuals with a basic cognitive capacity (just like animals) to be generally permissible. In addition, we also manipulated suffering capacity. We hypothesized that participants consider harming individuals with a high suffering capacity to be more morally wrong than individuals with a low suffering capacity.

Method

Participants

We recruited 812 US participants online via MTurk who received $0.5 payment for their participation. One hundred and eighty-seven were excluded for failing at least one attention check or manipulation check, leaving a final sample of 625 people (307 females; M_{age} = 35.72, SD = 11.30). We expected a small effect size. Our analysis showed that 787 participants were required to detect an effect size of f = .1 with an alpha of 0.05, and power of 0.80. We aimed to recruit 800 participants to account for any exclusions.

Design, materials, and procedure

The study had a 2 (cognitive capacity: basic vs. advanced) x 2 (suffering capacity: low vs. high) study design with both factors being manipulated between subjects. This study design and stimulus were loosely based on the study by Sytsma and Machery (2012). Participants read a fictional paragraph about a planet on a nearby solar system on which extraterrestrial individuals called 'Atlans' are the dominant form of life. Participants then read one of four descriptions of the Atlans depending on condition. The Atlans were described as either having suffering capacity ("feel pleasure and pain as well as other emotions such as fear, anger and sadness") or as having no suffering capacity ("don't feel

pleasure or pain or any other emotions such as fear, anger, or sadness"). Further, they were

described as either having an advanced cognitive capacity ("have opinions, beliefs, desires,

language, and culture. […] can make decisions for themselves. […] make plans for their

long-term future") or as having a basic cognitive capacity ("don't have opinions, beliefs,

desires, language, or culture, […] respond to their environment purely based on impulse and

instinct […] don't make plans for their long-term").

Following these descriptions, participants were given the vaccine scenario from Study

1, according to which ten healthy Atlans needed to be killed to develop a vaccine to save a

hundred other Atlans from dying of a disease. Next, participants responded to manipulation

check questions, which tested whether they correctly understood and remembered what

cognitive capacity level or suffering capacity level the extraterrestrials had. Finally,

participants completed the Speciesism Scale ($\alpha = .88$), the Oxford Utilitarianism Scale (OUS-

IH $\alpha = .78$; OUS-IB $\alpha = .73$), and answered demographic questions.

Results

A two-way ANOVA revealed that participants considered it more permissible to harm

ten individuals to save a hundred individuals when these individuals had a basic cognitive

capacity compared to if they had an advanced cognitive capacity, $F(1, 621) = 10.56$, $p = .001$,

$\eta_p^2 = .02$ (Figure 10). However, the level of suffering capacity made no difference on moral

judgements, $F(1, 621) = 0.34$, $p = 0.56$, $\eta_p^2 = .001$, nor was there an interaction between

cognitive capacity and suffering capacity $F(1, 621) = 0.06$, $p = .80$, $\eta_p^2 < .001$. In post-hoc

tests moral permissibility ratings for individuals without suffering capacity that either were

cognitively basic ($M = 4.44$, $SD = 1.71$) or cognitively advanced ($M = 3.95$, $SD = 1.83$) did

not differ statistically significantly, $p = .07$, $d = .28$. Similarly, in post-hoc tests moral

permissibility ratings for individuals with suffering capacity that either were cognitively basic

(M = 4.33, SD = 1.74) or cognitively advanced (M = 3.90, SD = 1.81) did not differ

statistically significantly, p = .14, d = .24.
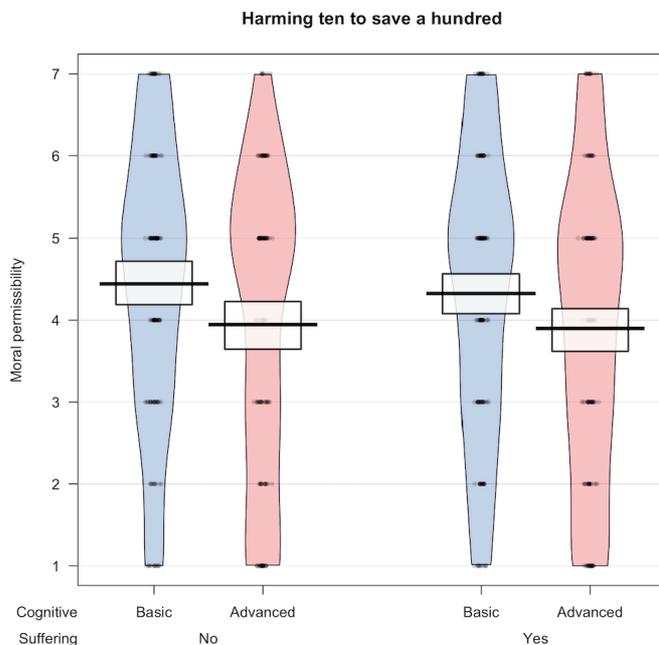
Figure 10. Moral permissibility of harming 10 individuals of a hypothetical species to saving

100 individuals of the same species, ranging from 1 (*Absolutely morally wrong*), to 4

(*Neither right nor wrong*), to 7 (*Absolutely morally right*). (Study 9)

Discussion

      The results of this study revealed that people hold stronger deontological constraints

against harming cognitively advanced beings compared to cognitively basic beings. This

suggests that the reason why people hold stronger deontological constraints against harming

animals than humans may be partly explained by the fact that people perceive animals to

have a lower cognitive capacity than humans. However, as we have seen in Study 8,

cognitive capacity cannot be the full explanation because people continue to have stronger

deontological constraints for humans even in cases where they have lower cognitive capacity than animals.

By contrast, the deontological constraints people hold for beings with higher suffering capacity were identical to the deontological constraints people hold for beings with lower suffering capacity. This is in line with Sytsma and Machery (2012) who also found no difference between people's judgments of harming sentient vs. non-sentient individuals to benefit humans, using a similar manipulation to the one used here. This lack of effect is surprising considering that in our previous studies we found that perceived suffering capacity predicted harm aversion and correlated with moral permissibility. It is also inconsistent with the finding that people are less likely to harm robots to save many humans if the robots are described as having high degree of affective experience than low degree (Nijssen, et al., 2019). One possibility is that our suffering capacity manipulation was suboptimal because it only focused on the ability to experience hedonic states and emotions, but not on an individual's level of consciousness. In Study 10 we address this by relying on an improved manipulation for suffering capacity.

Another limitation of this study is that participants may have found it hard to imagine extraterrestrial beings. The psychological distance for extraterrestrials may be higher than for terrestrial animals, which people know and may have even interacted with. It is therefore possible that people's thinking about animals and extraterrestrials fundamentally differs, which could make real world inferences from this study difficult.

Study 10: Suffering capacity

In Study 8, we found that people held stronger deontological constraints for humans than animals even when they perceived the humans to have a lower cognitive capacity than the animals. This showed that differences in perceived cognitive capacity cannot be the full

explanation of the effect. In Study 10, we turned to the related question of whether differences in perceived suffering capacity could explain the effect. In the previous studies we found that participants perceived animals to have a lower capacity to suffer than humans. It could, therefore, be that the weaker deontological constraints against harming animals could be completely explained by the fact that people perceive animals to have lower capacity to suffer. The capacity to suffer is widely regarded as a key driver of the aversion people have to harming others (Gray et al., 2012). (Note that even a utilitarian anti-speciesist would prioritize humans over animals if they believe animals suffer less than humans, cf. Singer, 1975). To test that hypothesis, we experimentally manipulated suffering capacity. More specifically, we presented participants either with a scenario that involved humans in a persistent vegetative state (PVS) who cannot experience any suffering but are still alive, or one that involved young dogs that were described as having extremely high capacity to suffer. As such, there were two conditions: puppies and patients.

Although the studies reported so far confirm the hypothesis that harm aversion plays a key role in explaining why people find it more permissible to harm animals compared to humans, it is plausible that differences in degree of harm aversion cannot fully explain the superior moral status that people ascribe to humans. A number of philosophers have argued that 'human' is used as a moral term and that to classify someone as human is already to regard them as meriting special treatment, *regardless* of what other capacities they may or may not possess (Diamond, 1978; Williams, 2009; for a critical approach to this tendency, see Singer, 1975). Indeed, many would be horrified by the idea that we could use living anencephalic human infants in scientific experiments or for organ donation, although such humans lack the capacity to feel pain or any other experience (McMahan, 2002). Our hypothesis was therefore that participants would continue to find it more permissible to harm a few puppies to save many puppies than to harm a few human patients to save many human

patients, despite being more averse to harm the puppies and believing that the patients have

lower suffering capacity than the puppies do.

Method

Participants

We recruited 203 US American participants online via MTurk who received $0.40

payment for their participation. Two participants were excluded for failing an attention

check, leaving a final sample of 201 people (93 females; M_{age} = 40.93, SD = 12.86). We

aimed for a sample size of 210 participants, as our a priori power analysis showed that 200

participants were required to detect a medium effect size of d = .4, with an alpha of 0.05, and

power of 0.80.

Design, materials and procedure

Participants were again presented with a modified version of the vaccine dilemma

from Study 1. In the puppies condition, participants were told that there are two dog shelters

that host many young puppies. The puppies were described as follows: "These puppies have

an extremely high capacity to experience happiness and pain. Usually they are very happy

and jump around all day long. But if they get harmed, they experience extremely strong

pain." In the patients conditions, participants were told that there are two clinics that host

PVS (persistent vegetative state) patients. The patients were described as follows: "These

patients have severe brain damage and have therefore permanently lost the capacity to

experience any happiness or pain. They will never feel anything, but they are still alive. This

means that if they get harmed, they will experience absolutely no pain." Next, as in prior

vaccine dilemmas, participants were asked whether they consider it morally right or wrong to

actively kill 10 unaffected puppies/PVS patients to save 100 other puppies/PVS patients.

On the next pages, participants were asked questions about the perceived suffering capacity ($\alpha = .98$) and harm aversion ($\alpha = .95$). The questions were identical to the ones asked in Study 4 with the only difference that in the three suffering capacity items point 1 of the response scale was "Not suffering at all", "No pain at all", and "Not feel hurt at all" to account for the fact that in this particular study the patients were described as having no suffering capacity at all. Finally, participants responded to the Speciesism Scale ($\alpha = .89$), Empathic Concern Scale ($\alpha = .94$), and demographic questions.

Results

Participants considered it more permissible to harm ten puppies in order to save 100 puppies ($M = 3.80$; $SD = 1.96$) than to harm 10 PVS patients to save 100 PVS patients ($M = 3.10$, $SD = 1.78$; Figure 11), $t(198) = 2.66$, $p = .008$, $d = .38$. Harm aversion for dogs ($M = 7.73$; $SD = 1.48$) was stronger than for patients ($M = 6.91$, $SD = 2.07$), $t(177) = 3.23$, $p = .001$, $d = .46$. Perceived suffering capacity for puppies ($M = 7.90$, $SD = 1.05$) was considerably higher than for patients ($M = 3.32$; $SD = 2.69$), $t(127) = 15.82$, $p < .001$, $d = 2.25$, which served as a manipulation check.
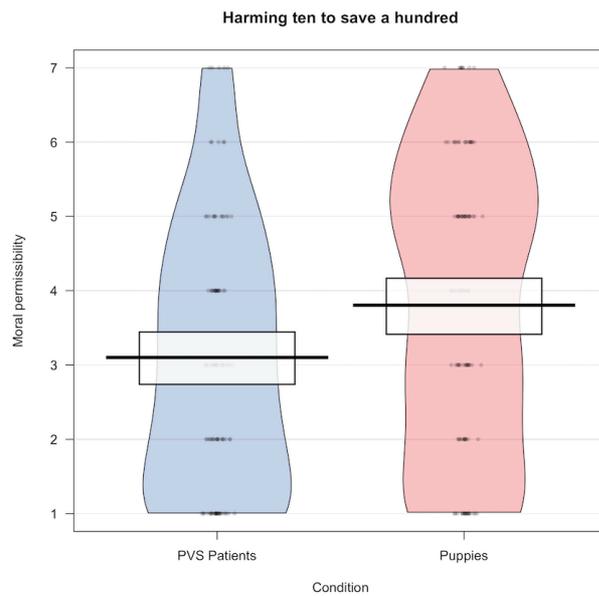
Figure 11. Moral permissibility of harming ten puppies to save a hundred puppies or of harming ten persistent vegetative state patients (described as having no suffering capacity) to save a hundred PVS patients, ranging from 1 (*Absolutely morally wrong*), over 4 (*Neither right nor wrong*), to 7 (*Absolutely morally right*). (Study 10)

Next, we tested the Harm Aversion Mediated Speciesism Model. The relationship between condition (patients vs. puppies) and moral permissibility was mediated by harm aversion. The bootstrapped indirect effect was (-.82)(-.29) = .23, 95% CI [.06, .45]. Thus, the indirect effect was statistically significant. However, note that in comparison to the other studies, the indirect effect was positive because participants were more averse to harm the puppies than the patients. We conducted a hierarchical regression to test which factors were driving harm aversion (Table 17). We found that speciesism did not moderate the effect of condition onto harm aversion. The bivariate correlation between speciesism and harm aversion was -.24 (*p* = .02) in the patients condition and -.34 (*p* < .001) in the puppies condition (Table 18). However, perceived suffering capacity significantly predicted harm

aversion. Thus, in line with our specific prediction for this study, the Harm Aversion

Mediated Speciesism model was only partly supported.

Finally, we found that the bivariate correlation between speciesism and suffering

capacity was stronger in the puppies condition ($r = -.36$, $p < .001$) than in the patients

condition ($r = -.15$, $p = .13$). Despite this, speciesism did not significantly moderate the effect

of condition onto suffering capacity ($\beta < .001$, $p = 1.0$), which is because differences in

suffering capacity were mostly explained by condition itself ($\beta = -.74$, $p < .001$).

Table 17

Hierarchical regression predicting harm aversion (Study 10)

Predictor measures	Step 1	Step 2
R^2	.11	.17
Condition	-.23	.06
Speciesism	-.28**	-.23*
Speciesism * Condition	.04	.04
Suffering		.40***

Note. Displaying standardized coefficients β. Condition was coded as 1 for PVS patients and as 0 for puppies.
†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

Table 18

Correlations between measures (Study 10)

	Permissibility	Speciesism	Harm av.	Suffering	EC
Speciesism	.27***				
Harm av.	-.22***	-.28***			
Suffering	.08	-.18**	.37***		

EC	-.17**	-.29***	.44***	.08
Condition	-.19**	.07	-.22***	-.75***	< .01

Note. Condition was coded as 1 for PVS patients and as 0 for puppies.
†*p* < .10. **p < .05. **p < .01. ***p < .001.

Discussion

The results of this study demonstrate that perceived suffering capacity alone cannot explain why people have weaker deontological constraints against harming animals than humans. Even in the extreme case we relied on in this study, in which the humans were described as having no suffering capacity at all and the animals were described as having extremely strong suffering capacity (and in addition having used an animal that people typically find particularly cute and likeable), participants consider it more permissible to harm animals than humans for the greater good. This suggests, in line with our hypothesis for this study, that there must be factors above and beyond perceptions of suffering capacity that contribute to the differences in strength of deontological constraints for animals and humans.

It is thus not surprising that the Harm Aversion Mediated Speciesism Model was only partly supported. In line with the model, suffering capacity predicted harm aversion such that participants were more averse to harming puppies than the humans (PVS patients) because they perceived the puppies to have greater suffering capacity. However, participants still considered it more permissible to harm a few puppies to save many than to harm a few PVS patients to save many despite having a greater general aversion to harm the puppies than the humans. This is in conflict with a model that assumes that moral permissibility is entirely driven by harm aversion.

Moreover, we found no statistically significant moderation of speciesism onto the relationship between condition and harm aversion. One possible explanation could be that—similar to Study 7—speciesism reduces the aversion to harm dogs (and in particular cute

puppies) less than the aversion to harm other animals. However, note that despite the non-significant moderation, the correlation between speciesism and aversion to harm puppies was stronger ($r = -.34$) than between speciesism and aversion to harm humans ($r = -.24$). Therefore, it is also possible that a moderation effect could be detected with a higher sample size. Either way, this study suggests that the speciesism moderation effect alone is unlikely to be the only driver of the effect.

Overall, these results suggest that further psychological factors can drive judgments about the permissibility of harm to humans and animals. In particular, the findings suggest that harm aversion is not the only mediator between species-membership and moral permissibility. One possibility is that speciesism could affect moral permissibility above and beyond the harm aversion path. For example, as mentioned in the introduction of this study, people may perceive humans, but not animals, to be deserving of special moral treatment irrespective of their capacity to suffer.

General Discussion

Across ten studies we found that people consider it more permissible to harm a few animals to save many animals than to harm a few humans to save many humans. The effect robustly showed in different types of dilemmas: in personal (Footbridge case) and impersonal (Sidetrack case) sacrificial dilemmas (Study 2), in both abstract (Study 4) and concrete (e.g., vaccination, Studies 1-3, 5-10) dilemmas, and even in the real-world context of opposition to medical experimentation (Study 6). While deontological constraints against harming humans were much stronger than those against harming animals, they differed slightly in strength for different animals. For example, deontological constraints against harming dogs were stronger than those against harming pigs (Study 1). Some people even showed deontological constraints for objects, albeit very weak ones (Study 5).

These findings closely fit the view we called Multi-level Weighted Deontology, on which there are deontological constraints against harming animals but these are considerably weaker than those enjoyed by humans. By contrast, our findings are hard to reconcile with the competing views we outlined in the introduction. Utilitarianism and Cross Species Deontology can be ruled out because people generally value humans more than animals. Kant's view can be ruled out because people do value animals at least to some extent (although this could also be due to instrumental reasons). Multi-level Uniform Deontology cannot accommodate the consistent finding that deontological constraints against harming animals are weaker than those against harming humans even if many of the same respective beings can be saved. Nozick's "utilitarianism for animals, Kantianism for people" is often thought to reflect commonsense intuitions (Kagan, 2019; Killoren & Streiffer, 2019). However, while being a good first approximation, Nozick's slogan does not account for the fact that people *do* have deontological constraints for animals and therefore are not completely "utilitarian" for animals. Furthermore, it does not account for the fact that deontological constraints can differ for different types of animals. Multi-level Weighted Deontology, in contrast, appropriately captures the finding that deontological constraints get weaker—without completely disappearing—the less people value the beings at stake. Although there was individual variation in the degree to which participants were willing to sacrifice some animals to save a greater number—with some participants demonstrating no inhibition about harming animals in such moral contexts—our data suggest that the majority of participants' responses is better described by Multi-level Weighted Deontology than by any of the other discussed moral views.

Underlying factors of the reduced deontological aversion to harm animals

We are not claiming that people explicitly endorse and follow the principles of Multi-level Weighted Deontology; it is unlikely that people explicitly grant deontological constraints to individuals and consciously weigh these constraints by the moral status they attribute to these individuals. Rather, our claim is that Multi-level Weighted Deontology is an accurate philosophical description of the pattern of moral judgments people make. This, however, leaves open the question of what psychological mechanisms underlie these judgment patterns.

In the introduction we hypothesized that people would consider it more permissible to harm animals than humans for the greater good because of a reduced general aversion to harm animals than humans (mediation from condition via harm aversion to moral permissibility). Further, we hypothesized that harm aversion is driven by perceived suffering capacity and perceived cognitive capacity and that harm aversion for animals is affected more by speciesism than harm aversion for humans (moderation). Across our studies we tested this Harm Aversion Mediated Speciesism Model and found support for it. However, that was not the case in all the studies, suggesting that further factors might be involved.

Harm aversion. Across our studies we found evidence that the weaker deontological constraints against harming animals than against harming humans are likely driven by a weaker aversion to harm animals in general. In Study 3, we systematically teased apart the deontological inclination to avoid harm from the utilitarian inclination to help as many as possible and found that the deontological inclination to avoid harm was much weaker for animals than for humans. In contrast, the utilitarian inclination did not statistically differ for animals and humans. In the majority of the studies in which we included the self-report harm aversion scales, we found that harm aversion significantly mediated the effect from condition (humans vs. animals) to moral permissibility.

In Study 10, we found that participants had stronger deontological constraints for patients in the persistent vegetative state than for puppies despite being more averse to harming the puppies. This suggests that harm aversion might not be the only driver of moral permissibility and that future research could expand our model to include further drivers of moral permissibility. One possibility is that people ascribe a special moral status to humans simply in virtue of them being human, regardless of their capacity for suffering or cognitive capacity. The philosopher Bernard Williams, for example, wrote that "creatures are treated in one way rather than another simply because they belong to a certain category, the human species. We do not… need to know any more about them." (Williams, 2006).

Setting this exception aside, our studies demonstrate that harm aversion is a strong predictor of the perceived moral permissibility of harming few to save a greater number. But what are the drivers of harm aversion? Since humans and animals differ in many ways, we investigated several possible factors that could explain why people are less averse to harming animals than humans.

Social Connectedness. In Study 7, we ruled out the possibility that the effect can be explained by the fact that people feel less socially connected to animals. We found that participants' deontological constraints against harming animals were still weaker than those against harming humans, even if participants felt more socially connected to the animals (e.g., their own dogs) than to the humans (e.g., complete strangers).

Suffering capacity. In all six studies in which we included the perceived suffering capacity scales, we found that perceived suffering capacity was a significant predictor of harm aversion. And since participants tended to believe that animals are less capable of suffering than humans (apart as expected from Study 10), this suggests that perceived suffering capacity could indeed partly explain why animals are granted weaker deontological constraints than humans. However, despite the important role of perceived suffering capacity,

our findings show that it cannot explain the effect fully. In Study 10, participants continued to have stronger deontological constraints for humans than animals even in cases where the humans have no suffering capacity at all (PVS patients) and the animals have an extremely strong suffering capacity.

Cognitive capacity. In Study 9, we found that participants held stronger deontological constraints against harming hypothetical beings with an advanced cognitive capacity than against harming hypothetical beings with a basic cognitive capacity. Since humans generally have a more advanced cognitive capacity than animals, this suggests that perceived cognitive capacity could partly explain the effect. This result is in line with the Kantian idea that a higher cognitive capacity—and in particular those aspects relating to rationality and autonomous agency—is ground for elevated moral status (even though people clearly reject the view, also associated with Kant, that animals have the same moral status as mere objects). However, although perceived cognitive capacity plays a role in driving judgments about the permissibility of harm, we found that it too cannot explain the effect fully. In Study 8, participants accepted stronger deontological constraints against harming humans who are severely cognitively impaired than against harming chimpanzees, even though participants believed that the chimpanzees had a more advanced cognitive capacity than the humans (including greater intelligence, rationality, ability to plan for the future, self-awareness, etc.).

Speciesism. Our studies also offer evidence that in addition to the contributions made by perceived suffering capacity and cognitive capacity, reduced aversion to harming animals than humans is also partly driven by differential moral value attribution based on species-membership itself, i.e., speciesism. According to this hypothesis, people are less averse to harm animals than humans simply because animals are animals and humans are humans. There are three convergent types of evidence supporting this hypothesis.

First, in four of the six studies in which we measured harm aversion, we found that speciesism as measured by the Speciesism Scale (Caviola et al., 2019) significantly moderated the effect of condition (humans vs. animals) onto harm aversion. In the two studies (Study 7 and 10) in which the moderation was not significant the animal species were dogs. It is therefore possible that the lack of significant moderation effects in these studies was specific to the use of dogs, since speciesism appears to be a weaker predictor for aversion to harming dogs than aversion to harming other animals. Nevertheless, the lack of significant moderation in these two studies highlights that speciesism is not the only predictor of the reduced aversion to harm animals.

Second, in the process dissociation study (Study 3) individual differences in speciesism correlated strongly ($r = -.65$) with the deontological inclination against harming animals. Speciesism did not correlate with the deontological inclination against harming humans, and neither did it correlate with the utilitarian inclination to help as many humans or animals as possible. The process dissociation approach allowed us to disentangle the deontological and utilitarian inclinations from each other, which usually both contribute to the overall moral judgment in sacrificial moral dilemmas (cf. Conway, & Gawronski, 2013).

Third, throughout our studies we found that deontological constraints against harming animals remained weaker than those against harming humans even when other plausible drivers, such as perceived suffering capacity, ascribed cognitive capacity, and social connectedness were accounted for. Humans and animals differ on multiple dimensions and it is possible that there are other factors that we did not consider. However, the fact that a robust effect remains even after controlling for key factors suggested both by psychological research and ethical debate about moral status provides probabilistic—if not conclusive— evidence for the speciesism hypothesis.

Furthermore, we found evidence that speciesism is associated with perceptions of reduced suffering capacity and reduced cognitive capacity of animals. This is in line with previous research showing that people tend to de-mentalize animals (Bastian et al., 2012). It suggests that speciesism may not only reduce aversion to harm animals directly but also indirectly via attributing lower mental capacities to them, which in turn could reduce harm aversion.

Our studies thus offer a degree of empirical support to the influential (if controversial) philosophical argument claiming that the common view that humans are morally more important than animals is based in speciesism (McMahan, 2002; Singer, 1993). However, future research is called to explore in more detail why people have a greater willingness to harm animals than humans, even in cases where the humans suffer less and are less cognitively capable than the animals.

In sum, our studies mostly confirm the Harm Aversion Mediated Speciesism Model but also show that the model needs to be extended: People consider it more permissible to harm animals for the greater good than to humans for the greater because they have generally a weaker aversion to harm animals than humans. The weaker aversion to harm animals is driven in part by people believing that animals suffer less, and that they have a lower cognitive capacity, as well as by speciesism. It is possible that additional drivers of moral permissibility exist that we did not explore in our studies.

Thus, the two models we have discussed in this paper—Harm Aversion Mediated Speciesism and Multi-Level Weighted Deontology—both describe the moral judgments people make in our studies. While Multi-Level Weighted Deontology is a philosophical description of these judgments and is contrasted with alternative philosophical views, Harm Aversion Mediated Speciesism describes the psychological mechanisms of these judgments in more detail. They are two sides of the same coin.

Moral judgments in cross-species sacrifice cases

So far, we have only focused on the type of individuals that have to be harmed but not those that can be helped. However, in the cross-species sacrifice cases of Study 4, we found that people also take into account what type of individuals can be saved. We found that people considered it only slightly more permissible to harm pigs to save many humans than to save many pigs. In contrast, people were much more sensitive to the type of individuals that could be saved if humans had to be harmed. They were strongly opposed to harming humans in order to save many pigs, whereas harming humans to save many humans was considered to be less wrong on average. How can this asymmetry be explained?

One way to explain this is as follows: The lower the moral status of the individual that needs to be harmed, the weaker the deontological constraint against harming that individual. The stronger the deontological constraints, the more difficult it is to overrule them. People have a rough threshold—that varies between individuals and contexts—for when to overrule a given deontological constraint. Whether this threshold is met depends on the utilitarian cost-benefit analysis, which people will engage in if they are capable of doing so. In cases where animals have to be harmed to save either many animals or many humans, the threshold is clearly met in both cases—though since people value humans much more than animals, it is more easily met if humans instead of animals can be saved. In contrast, however, in cases where humans have to be harmed to save either many humans or many animals, the threshold is not clearly met, and the difference in how easily it is met is far greater than in the previous cases. The threshold for harming humans to save many humans may be met to some extent in some people. But since people value humans much more than animals, the threshold for harming humans to save animals will be met, if at all, only if a *vast* number of animals will

be saved. As a consequence, the extent to which deontological constraints against harming are overruled differs across these four cases and results in the observed asymmetry.

This explanation assumes that people assign a certain value to the life of a single animal and another value to the life of a single human—all using the same moral value "currency". These values are then aggregated each based on the number of individuals on each side and contrasted against each other. The contrasted aggregated values are then matched against a common threshold value that is measured in this shared currency. An advantage of this explanation is that it could also deal with mixed sacrifice cases, such as a case where 55 pigs have to be harmed to save 50 pigs plus 2 humans, which is a type of case we did not test but future research could look into.

A reason to be skeptical of the described view could be that it appears psychologically implausible that people engage in such complex calculations and rely on a common moral value currency. Instead, it may be more plausible that a much rougher estimate, that is itself partly affect-based, is at work. One possibility is that the threshold for when deontological constraints should be overruled is different depending on what type of individuals can be saved. The threshold is lower, the lower the perceived moral status of the individuals that can be saved. This view, however, is harder to reconcile with the idea of a fixed aversive response against harming dictating the strength of the deontological constraint since it suggests that the constraint against harming does not arise from an affective response to considerations of the harmful act in isolation, but is instead modulated by which individuals would be saved by that act, and how many. In other words, on this view the degree of aversion to harm and the calculation of consequences interact rather than being generated independently, as is often assumed by current models of moral judgement.

Implications for the psychology of moral judgment

Current psychological research on deontological rules against harming others often models such rules on Kant's ethics, and therefore understands them as representing absolute prohibitions against certain actions. It is implausible, however, that lay people regard deontological rules as absolute in this way (Kahane, 2015). Lay people's intuitions accept a plurality of moral rules which can conflict in some cases. When this happens one rule will typically overrule another (e.g., we may break a minor promise if keeping it would be deeply unfair). Similarly, deontological constraints can be overruled when the consequences of following them are too severe—e.g., lying when this will prevent murder and, more controversially, the use of 'enhanced interrogation methods' to prevent terrorist attacks. The present research provides a demonstration of this central feature of commonsense morality. While the absolutist understanding of deontological rules suggest the all-or-nothing picture suggested by Nozick's proposal—absolute prohibitions protecting humans, cost-benefit analysis for animals—our findings strongly support a Multi-level Weighted Deontology account of judgments about harm towards humans and animals. We repeatedly found that deontological constraints are a matter of degree. And when and whether they will be overruled is a function both of the individuals to be harmed, the individuals to be saved, and the numbers of each.

Our findings consequently also show that moral thinking about harm to humans and animals is not fundamentally different but rather varies in degree. People value animals much less than humans and accordingly grant them much weaker deontological constraints. Importantly, though it is not the case that people have no deontological constraints for animals whatsoever. This difference in degree, however, can be very substantial, and most people consider it fairly wrong to harm humans while fairly permissible to harm animals for the greater good.

Animals are seen neither as objects nor as persons

Our findings support Nozick's suggestion (1974) that people place animals into a moral category between humans and objects. In some respects, animals are perceived as having a moral status similar to that of inanimate objects: like objects, people consider it morally permissible to own animals, to treat them as a tool for another's purpose, to deny them of their autonomy, or to treat them as interchangeable with other animals—properties that the philosopher Martha Nussbaum (1995) has identified as aspects of objectification. However, in contrast to objects, animals are seen as sentient beings that deserve *some* moral protection. For example, we found that people consider it permissible to destroy an object but not to harm an animal, if the sole purpose is to fulfil a personal preference. (The case of meat consumption appears to be an exception and might be related to people's tendency to de-mentalize farmed animals; see Bastian et al., 2012)

The fact that deontological constraints for humans are much higher than those for animals is reflected by the fact that humans are seen as possessing inalienable rights, but such rights are rarely, if ever, ascribed to animals. Most people of Western societies are firm believers in basic human rights such as a right to life, bodily integrity and autonomy—rights that are inalienable and absolute. Even humans without advanced cognitive capacity such as human infants or cognitively severely disabled humans are granted these same basic rights. Animals, in contrast, are not granted equivalent rights, not even the most basic ones. This shows, for example, in the context of medical experimentation or exploitation for consumption.

The strength of deontological protections people grant animals may predict how well they treat animals more generally. Historically, moral attitudes towards animals have been changing (Kelch, 2012; Pinker, 2012). In medieval Europe, for example, burning cats was

considered a form of entertainment (Benton, 1997). Today, most people are opposed to unnecessary cruelty to animals (Vaughn et al., 2009). And since the 1970s, the animal rights movement has emerged, advocating not just for better treatment of animals but also for granting legal rights to animals (Singer, 1973). In our studies, we found that people had much weaker constraints against harming animals than against harming humans. But they still (on average) granted animals at least some weak deontological protections. It is possible that a few centuries ago people would have granted animals even weaker deontological protections than today, and that in the future they will grant them even stronger ones. That being said, the deontological protections that people currently grant animals are not yet strong enough to make people view it wrong to harm animals in order to benefit humans.

Limitations and future research

We have identified Multi-level Weighted Deontology as the model that best describes people's intuitions about when it is permissible to harm humans and animals. Future research could explore the model further and make it more precise. Future research could look at a wider range of cross-species sacrifice cases as well as at mixed cases in which a group of individuals consisting of different species would have to be harmed or saved.

We concluded that there are likely multiple factors that explain why aversion to harm animals is weaker than the aversion to harm humans, amongst them perceived suffering capacity, perceived cognitive capacity, and speciesism. However, our studies cannot rule out other underlying factors that we have not explicitly explored in our studies—such as the appearance and behavior of the respective individuals or the fact that people may be more familiar with thinking about moral rights for humans than animals. Future research could also attempt to estimate the relative degree to which each factor contributes to the effect.

In Study 10 we found evidence that species-membership can affect permissibility of harming a few to save many without being mediated by harm aversion. Future research could attempt to replicate this finding and explore this potential additional path in more depth. For example, it could be tested whether people grant humans special moral status simply in virtue of them being human, regardless of their capacities.

While we currently consider it plausible that speciesism is indeed a driving factor—echoing an influential philosophical view (Singer, 1975)—it is possible that the effect can instead be explained by other factors that are typically associated with species-membership. Future research could investigate which factors make people conceptualize a being as an animal or as a human. For example, would people conceptualize a cognitively highly advanced chimpanzee still as an animal or, instead, as a person and thus, perhaps no longer as an animal? What if an animal closely resembled humans in appearance and aspects of behavior while lacking normal human cognitive capacity? An empirical investigation of these questions could give us a deeper understanding of the psychological mechanisms at play.

Conclusion

Bentham argued that what matters is whether a being suffers, not how smart it is, and utilitarians hold that we should maximize utility—by saving the greater number—regardless of species-membership. Kant held that there are some things we must never do to other rational beings, even if they would maximize utility, and that people therefore enjoy deontological protections from certain kinds of trade-offs. Nozick proposed a way in which these two views might be combined: that Kant is right about humans, but Bentham is right about animals: "Utilitarianism for Animals, Kantianism for People". We found that neither of these views succeeds in capturing how most people think about harm to humans vs. to animals. People are deontological all the way down—but they do not regard deontological

protections as absolute but as getting increasingly weaker as we go down the 'chain of

being'.

References

Amiot, C. E., & Bastian, B. (2015). Toward a psychology of human-animal relations. *Psychological Bulletin*, *141*(1), 6–47. https://doi.org/10.1037/a0038147

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., … Rahwan, I. (2018). The Moral Machine experiment. *Nature*, *563*(7729), 59. https://doi.org/10.1038/s41586-018-0637-6

Bastian, B., Loughnan, S., Haslam, N., & Radke, H. (2012). Don't Mind Meat? The Denial of Mind to Animals Used for Human Consumption. *Personality & Social Psychology Bulletin*, *38*, 247–256. https://doi.org/10.1177/0146167211424291

Bauman, C. W., McGraw, A. P., Bartels, D. M., & Warren, C. (2014). Revisiting External Validity: Concerns about Trolley Problems and Other Sacrificial Dilemmas in Moral Psychology. *Social and Personality Psychology Compass*, *8*(9), 536–554. https://doi.org/10.1111/spc3.12131

Bentham, J. (1780). 1988 The principles of morals and legislation. *Amherst, NY: Prometheus Books*.

Benton, J. R. (1997). *Holy terrors: Gargoyles on medieval buildings*. Abbeville Press New York, NY.

Bostyn, D. H., Sevenhant, S., & Roets, A. (2018). Of mice, men, and trolleys: Hypothetical judgment versus real-life behavior in trolley-style moral dilemmas. *Psychological Science*, *29*(7), 1084–1093.

Caviola, L., & Capraro, V. (2019). Liking but devaluing animals: emotional and deliberative paths to speciesism. *Social Psychological and Personality Science*. https://doi.org/10.1177/1948550619893959

Caviola, L., Everett, J. A. C., & Faber, N. S. (2019). The moral standing of animals: Towards

      a psychology of speciesism. *Journal of Personality and Social Psychology*.

      https://doi.org/10.1037/pspp0000182

Caviola, L., Kahane, G., Everett, J. A. C., Teperman, E., Savulescu, J., & Faber, N. S. (2020,

      August 6). Utilitarianism for animals, Kantianism for people? Harming animals and

      humans for the greater good. Retrieved from osf.io/nt69s/

Caviola, L., Schubert, S., Kahane, G., & Faber, N. S. (2020). Why people prioritize humans

      over animals: A framework for moral anthropocentrism. *Manuscript under Review*.

Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral

      decision making: A process dissociation approach. *Journal of Personality and Social*

      Psychology, *104*(2), 216–235. https://doi.org/10.1037/a0031021

Crimston, D., Bain, P. G., Hornsey, M. J., & Bastian, B. (2016). Moral expansiveness:

      Examining variability in the extension of the moral world. *Journal of Personality and*

      Social Psychology, *111*(4), 636.

Crockett, M. J. (2013). Models of morality. *Trends in Cognitive Sciences*, *17*(8), 363–366.

Goodwin, G. P. (2015). Experimental approaches to moral standing. *Philosophy Compass,*

      10(12), 914-926.

Cushman, F. (2013). Action, outcome, and value: A dual-system framework for morality.

      Personality and Social Psychology Review, *17*(3), 273–292.

Dhont, K., Hodson, G., & Leite, A. C. (2016). Common Ideological Roots of Speciesism and

      Generalized Ethnic Prejudice: The Social Dominance Human–Animal Relations

      Model (SD-HARM). *European Journal of Personality*, *30*(6), 507–522.

      https://doi.org/10.1002/per.2069

Diamond, C. (1978). Eating meat and eating people. *Philosophy, 53*(206), 465-479.

Everett, J. A., Caviola, L., Savulescu, J., & Faber, N. (2018). Speciesism, generalized

　　　prejudice and perceptions of prejudiced others. *Group Processes and Intergroup*

　　　Relations.

Everett, J. A., Faber, N. S., Savulescu, J., & Crockett, M. J. (2018). The costs of being

　　　consequentialist: Social inference from instrumental harm and impartial beneficence.

　　　Journal of Experimental Social Psychology, *79*, 200–216.

Everett, J. A., Pizarro, D. A., & Crockett, M. J. (2016). Inference of trustworthiness from

　　　intuitive moral judgments. *Journal of Experimental Psychology: General*, *145*(6),

　　　772.

Francione, G. L. (1995). *Animals Property & The Law*. Retrieved from

　　　https://www.jstor.org/stable/j.ctt1bw1jm9

Gennet, E., & Altavilla, A. (2016). Paediatric research under the new EU regulation on

　　　clinical trials: Old issues new challenges. *European Journal of Health Law*, *23*(4),

　　　325–349.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*,

　　　315(5812), 619–619.

Gray, K., Young, L., & Waytz, A. (2012). Mind Perception Is the Essence of Morality.

　　　Psychological Inquiry, *23*(2), 101. https://doi.org/10.1080/1047840X.2012.651387

Greene, J. (2009). The cognitive neuroscience of moral judgment. *The Cognitive*

　　　Neurosciences, *4*, 1–48.

Greene, J. (2014). *Moral tribes: Emotion, reason, and the gap between us and them*. Penguin.

Gruen, L, (2011). The Moral Status of Animals. The Stanford Encyclopedia of Philosophy

　　　(Fall 2017 Edition)

Hester, N., & Gray, K. (2019). *The Moral Psychology of Raceless Genderless Strangers*.

Holyoak, K. J., & Powell, D. (2016). Deontological coherence: A framework for commonsense moral reasoning. *Psychological Bulletin*, *142*(11), 1179.

Horta, O. (2010). What is Speciesism? *Journal of Agricultural and Environmental Ethics*, *23*(3), 243–266. https://doi.org/10.1007/s10806-009-9205-2

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513–541.

Kagan, S. (2019). *How to Count Animals, more or less*. Oxford University Press.

Kahane, G. (2015). Sidetracked by trolleys: Why sacrificial moral dilemmas tell us little (or nothing) about utilitarian judgment. *Social Neuroscience*, *10*(5), 551–560.

Kahane, G., Everett, J. A. C., Earp, B. D., Caviola, L., Faber, N. S., Crockett, M. J., & Savulescu, J. (2018). Beyond sacrificial harm: A two-dimensional model of utilitarian psychology. *Psychological Review*, *125*(2), 131–164. https://doi.org/10.1037/rev0000093

Kant, I. (1785). *Groundwork of the Metaphysics of Morals*.

Kelch, T. G. (2012). A Short History of (Mostly) Western Animal Law: Part I. *Animal L.*, *19*, 23.

Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M., & Altman, D. G. (2010). Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research. *PLoS Biology*, *8*(6), e1000412.

Killoren, D., & Streiffer, R. (2019). Utilitarianism about animals and the moral significance of use. *Philosophical Studies*, 1–21.

Kurzban, R., DeScioli, P., & Fein, D. (2012). Hamilton vs. Kant: Pitting adaptations for altruism against adaptations for moral judgment. *Evolution and Human Behavior*, *33*(4), 323–333. https://doi.org/10.1016/j.evolhumbehav.2011.11.002

McMahan, J. (2002). *The ethics of killing: Problems at the margins of life*. Oxford University
    Press, USA.

Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in
    Cognitive Sciences*, *11*(4), 143–152.

Muda, R., Niszczota, P., Bia\lek, M., & Conway, P. (2018). Reading dilemmas in a foreign
    language reduces both deontological and utilitarian response tendencies. *Journal of
    Experimental Psychology: Learning, Memory, and Cognition*, *44*(2), 321.

Nichols, S., & Mallon, R. (2006). Moral dilemmas and moral rules. *Cognition*, *100*(3), 530–
    542.

Nijssen, S. R., Müller, B. C., Baaren, R. B. van, & Paulus, M. (2019). Saving the Robot or
    the Human? Robots Who Feel Deserve Moral Care. *Social Cognition*, *37*(1), 41–S2.

Nozick, R. (1974). Constraints and Animals. In *Anarchy, State and Utopia* (pp. 35–42). New
    York: Basic Books.

Nuffield Council on Bioethics (Ed.). (2005). *The ethics of research involving animals*.
    London: Nuffield Council on Bioethics.

Nussbaum, M. C. (1995). Objectification. *Philosophy & Public Affairs*, *24*(4), 249–291.

Petrinovich, L., O'neill, P., & Jorgensen, M. (1993). An empirical study of moral intuitions:
    Toward an evolutionary ethics. *Journal of Personality and Social Psychology*, *64*(3),
    467.

Pinker, S. (2012). *The better angels of our nature: Why violence has declined*. Penguin
    Group USA.

Regan, T. (1987). The case for animal rights. In *Advances in animal welfare science 1986/87*
    (pp. 179–189). Springer.

Regan, T., & Singer, P. (1989). *Animal rights and human obligations*.

Schubert, S., Caviola, L., & Faber, N. S. (2019). the psychology of existential Risk: Moral Judgments about Human extinction. *Scientific Reports, 9*(1), 1-8.

Singer, P. (1975). *Animal liberation: A new ethic for our treatment of animals*. New York: HarperCollins.

Singer, P. (1993). Practical Ethics, Cambridge University Press.

Starmans, C., & Friedman, O. (2016). If I am free, you can't own me: Autonomy makes entities less ownable. *Cognition*, *148*, 145–153. https://doi.org/10.1016/j.cognition.2015.11.001

Sytsma, J., & Machery, E. (2012). The Two Sources of Moral Standing. *Review of Philosophy and Psychology*, *3*(3), 303–324. https://doi.org/10.1007/s13164-012-0102-7

Thomson, J. J. (1984). The trolley problem. *Yale LJ*, *94*, 1395.

Topolski, R., Weaver, J. N., Martin, Z., & McCoy, J. (2013). Choosing between the emotional dog and the rational pal: A moral dilemma with a tail. *Anthrozoös*, *26*(2), 253–263.

Trémolière, B., & Bonnefon, J. F. (2014). Efficient kill–save ratios ease up the cognitive demands on counterintuitive moral utilitarianism. *Personality and Social Psychology Bulletin, 40*(7), 923-930.

Varner, G. E. (1994). The prospects for consensus and convergence in the animal rights debate. *Hastings Center Report*, *24*(1), 24–28.

Vaughn, M. G., Fu, Q., DeLisi, M., Beaver, K. M., Perron, B. E., Terrell, K., & Howard, M. O. (2009). Correlates of cruelty to animals in the United States: Results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Journal of Psychiatric Research*, *43*(15), 1213–1218.

Wiech, K., Kahane, G., Shackel, N., Farias, M., Savulescu, J., & Tracey, I. (2013). Cold or

    calculating? Reduced activity in the subgenual cingulate cortex reflects decreased

    emotional aversion to harming in counterintuitive utilitarian judgment. *Cognition,*

    126(3), 364-372.

Williams, B. (2009). *The human prejudice. Peter Singer Under Fire: The Moral Iconoclast*

    Faces His Critics, *3*, 77.