Hybrid theories, psychological plausibility, and the human/animal divide

Bob Fischer[1] · Clare Palmer[2] · T. J. Kasperbauer[3]

Abstract A hybrid theory is any moral theory according to which different classes of individuals ought to be treated according to different principles. We argue that some hybrid theories are able to meet standards of psychological plausibility, by which we mean that it's feasible for ordinary human beings to understand and act in accord with them. Insofar as psychological plausibility is a theoretical virtue, then, such hybrid theories deserve more serious consideration. To make the case for this view, we explain what psychological plausibility is and why we might value it, why the human/animal divide appears to be an entrenched feature of human psychology, and why Robert Nozick's hybrid theory doesn't go far enough. Finally, we make the case that a more promising psychologically plausible hybrid theory, with respect to humans and animals, will be (at least) at tribrid theory—that is, positing three domains rather than two.

Keywords Hybrid theories · Robert Nozick · Animal ethics · Psychological plausibility

✉ Bob Fischer
fischer@txstate.edu

Clare Palmer
c.palmer@tamu.edu

T. J. Kasperbauer
tjkasperbauer@gmail.com

[1] Texas State University, San Marcos, USA

[2] Texas A&M University, College Station, USA

[3] Indiana University Center for Bioethics, Indianapolis, USA

1 Introduction: hybrid theories

Following the convention in this special issue, let's say that a hybrid theory is any moral theory according to which different classes of individuals ought to be treated according to different principles. A uniform theory, by contrast, is one according to which a single principle should be applied across the board.[1] Perhaps the most famous hybrid theory—or, at least, the most famous slogan for a hybrid theory—was suggested by Robert Nozick (1974). According to Nozick's "Utilitarianism for animals, Kantianism for people," we should "maximize the total happiness of all living beings [while respecting] stringent side constraints on what [we] may do to human beings."

But this isn't the only hybrid theory available to us: there are as many hybrid theories as there are combinations of moral theories and domains. Obviously, most of these hybrid theories have very little going for them. For example, "Scanlonian contractualism for white men, consequentialist ecocentrism for everything else" isn't particularly promising. Still, *some* hybrid theories may be worth exploring insofar as we value *psychological plausibility*—i.e., and very roughly, insofar as it's a virtue of a moral theory that it's feasible for ordinary human beings to understand and act in accord with that theory. Of course, that's a long way from an argument for hybrid theories generally, much less any particular hybrid theory. Still, it would give us reason to give such theories some consideration.

Here's the basic idea. Let's limit our attention to hybrid theories structured around the human/animal divide. Within this group of hybrid theories, we suggest that there's an important distinction. Some hybrid theories base the different treatment owed to different classes of individuals on capacities or relationships those individuals have. Consider, for instance, a relative of Nozick's "Utilitarianism for animals, Kantianism for people": "Utilitarianism for the merely sentient, Kantianism for agents." Whether or not the contrast between being "merely sentient" and an agent can fully justify the application of different principles that this theory involves, this theory is clearly trying to explain the difference in a traditional way: namely, by appealing to the characteristics of the individuals to whom the principles apply.

In principle, though, hybrid theories don't need to give these kinds of reasons for different normative treatment of different classes. Instead, they might appeal to features of the individuals who have to *use* the theory—the individuals who are supposed to act in accord with its dictates—rather than the beings to whom the theory applies. As noted above, we're specifically interested here in hybrid theories

---

[1] As this definition suggests, hybrid theories are hybrid with respect to normative principles, not necessarily the values they recognize as fundamental. A form of consequentialism that says we ought to maximize both well-being and justice is committed to value pluralism, but it isn't a hybrid theory in our sense of that term: there is only one normative principle. We'll contrast hybrid theories with uniform theories (a single normative principle), rather than monist theories (a single fundamental value), to keep the theory / value distinction clear in what follows. Finally, we should note that some people also talk about hybrid theories in the context of role ethics, defending adoption of different normative principles based on the role you're currently occupying; see, e.g., Stone (1988). We don't discuss that conception here.

that appeal to *psychological plausibility* to explain which classes of beings should be treated according to which principles. If psychological plausibility is a theoretical virtue, then there is no guarantee that the classes of individuals that the theory highlights will map precisely onto classes defined by properties that have long been taken to be morally relevant, such as "having the capacity for rational thought" or "being sentient." Depending on the characteristics of the individuals who use the theory, some mismatch may occur.

It is, of course, widely held among animal ethicists that species membership is morally irrelevant; being human per se makes no moral difference. And on one level, nothing we say here challenges that: we offer no defense of the view that being human is, directly and in itself, a morally relevant property. However, if psychological plausibility is a theoretical virtue, and if taking account of the human/animal divide is sufficiently important for achieving psychological plausibility, then being human could be morally relevant *indirectly*—i.e., via the psychologies of those who use the theory. Finally, if this means that different principles should be applied to different classes of individuals, we could get the beginnings of an argument for a hybrid view.

Our aim here is to explore this idea—to sketch the moves that would be required—not to defend it against all comers. The sketching is work enough given the space available. So, we're not arguing that we ought to accept, much less believe, any particular hybrid theory. Nor are we arguing that, all things considered, we ought to prefer hybrid theories over uniform theories. Instead, we're only arguing for a weak conditional claim: *if* it's a virtue of a moral theory that it's psychologically plausible, then hybrid theories, such as the one we briefly outline, should at least be given a hearing.

We'll begin by discussing psychological plausibility, explaining why we think we should care about it and replying to some obvious worries. After making these general points, we'll consider why some hybrid theories might be psychologically plausible. We'll argue that there are good reasons to think that such theories can recognize a human/animal divide, since research suggests that it's one of the "deep features" of human psychology, making it particularly difficult to achieve dramatic changes in the way humans relate to animals, at least for the foreseeable future.[2] As we argue, though, the human/animal divide doesn't appear to be enough; there are other contours to human/animal relations that also appear to be worth considering.

---

[2] Nozick may have recognized this in suggesting his hybrid view, but either way, and as we'll discuss, we don't think that Nozick's hybrid view goes far enough in terms of psychological plausibility. This is no criticism of Nozick, as he didn't actually endorse utilitarianism for animals, Kantianism for people. He dismissed the view as inadequate almost as soon as he introduced it.

2 What is psychological plausibility?

Let's begin with a bit more detail about psychological plausibility. At the most general level, moral theories are psychologically plausible insofar as they respect the capacity of human beings to understand and act on them. Unsurprisingly, a great deal turns on how we flesh out the key notions here, leading to very different accounts of psychological plausibility and associated stories about why we should value it. Though it usually isn't framed that way, we can think of the "ought implies can" principle as the weakest form of psychological plausibility, which requires that it be possible, at least in principle, for individuals to do what the theory says they ought to do. In that vein, Owen Flanagan (1991) argued that moral theories should satisfy the "Principle of Minimal Psychological Plausibility." That is, they ought to "[m]ake sure when constructing a moral theory or projecting a moral ideal that the character, decision processing, and behavior prescribed are possible, or are perceived to be possible, for creatures like us" (32). But since he only requires that "the recommended ideals be possible under some conceivable social arrangement or other" (201), this ends up being a low bar. After all, many "conceivable social arrangements" are vastly different from anything observed in current human societies, and it's unclear why we should consider them all minimally plausible.

The more interesting versions of psychological plausibility place more substantive constraints on moral theories, as found in discussions of non-ideal theory in political philosophy (Hamlin & Stemplowska, 2012; Valentini, 2012). Non-ideal theory explicitly aims to take into account the feasibility of moral and political ideals, given real-world constraints. Pursuing social and institutional change requires that we account for "the actual causal processes that limit the range of feasible alternatives" (Wiens, 2013 p. 326), and human psychology is a prominent causal constraint on the social arrangements we can pursue. In particular, human psychology can create constraints on moral change that aren't amenable to modification over time. Non-ideal theory is motivated by a relatively modest understanding of what people can do, combined with an "ought implies can" argument against having certain moral and political objectives that conflict with the relevant practical constraints. It's a mistake, on this view, to advocate for moral changes that aren't realizable—working with a modest sense of "realizable"—given the world in which we find ourselves.

Drawing from these accounts, Kasperbauer (2017b) proposes three more demanding criteria to build on Flanagan's minimal criteria for psychological plausibility:

1. *Impact*: An ethical theory is psychologically plausible insofar as the ideas it promotes are capable of having the intended psychological effect on currently existing human beings.
2. *Achievability*: An ethical theory is psychologically plausible insofar as it promotes ideas and actions that are achievable by human beings, if not now then at some point in the future.
3. *Transition*: If an ethical theory requires currently existing psychological profiles to undergo significant modification, that ethical theory is psychologically

plausible insofar as it can explain how the psychological transition could occur, given current understandings of human psychology (p. 125).

In other words, a theory is psychologically plausible insofar as it's (1) currently adoptable, (2) realistic in terms of its demands for long-term psychological change, and (3) there's a coherent psychological story about how we can get from non-adoptability and unrealistic demands to the proposed state. In contrast with the minimal view that Flanagan proposes, let's say theories that meet all three conditions are *robustly* psychologically plausible. Indeed, one way of thinking about a standard objection to leading ethical theories—namely, that they are overly demanding—is that they lack robust psychological plausibility in this sense. That is, a theory's being overly demanding means that its demands are too strenuous for it to be widely adopted now; and, since there's no obvious route by which people might change to tolerate such strenuous demands in the future, it fails to respect central features of human psychology. When we talk about psychological plausibility in what follows, it's this robust variety in which we're interested.

3 Why psychological plausibility matters

Of course, some reject the idea that moral theories should be constrained by psychological plausibility. If a theory is correct, but most people refuse to accept or act on it, then that's due to their moral failing or moral weakness, rather than a problem with the theory itself. Though we are sympathetic to such a position, we think that even the most ardent idealists should value psychological plausibility in moral theorizing.

First, many moral theories aim to provide guidance for ordinary human beings, and guidance requires that there are human beings properly situated to follow the theory's demands. The different accounts of psychological plausibility outlined above all essentially ask how a moral theory could possibly provide guidance for the sorts of creatures we are. Of course, human beings change over time, so we must ask what we are justified in believing that human beings could possibly become in order to meet a theory's prescriptions. There is also inevitably some uncertainty in determining how human beings might change in the future (especially given variation across cultures). Nonetheless, it would seem to count against a theory if none of the vast documentation of human behavior, from history, anthropology, psychology, or the other social sciences, can support the plausibility of the theory's prescriptions. Absent a reasonable conception of how a theory might cohere with human psychology, either now or in the future, it's not clear why we would think the theory can provide guidance for ordinary human beings.

It is important, however, not to be overly cautious about the possibility of human change. Someone might have thought, at certain moments in history, that psychological plausibility requires accepting various beliefs and attitudes that underpin racism, sexism, and xenophobia. However, that thought would have been mistaken. Taking human biases into account in moral theorizing is different from

endorsing or supporting those biases. Moral attitudes have obviously changed throughout human history, and we should expect those changes to continue. But clearly some routes of change are easier than others (as illustrated by the history of progress on social justice issues). We can acknowledge the existence of entrenched biases without also being overly concessive to the status quo—a point to which we'll return later.

The second reason to value psychological plausibility is that many ethicists working on issues pertaining to animals have specific goals for social change. They want moral theories that are both true and that can improve the lives of animals. Making meaningful progress toward those goals may well be easier with a psychologically plausible theory. Consider, for example, the goal of improving animal welfare (in any context). It surely matters, for the purposes of achieving such a goal, how people in general currently use animals and whether they support improving animal welfare, not to mention whether their conception of "welfare" is compatible with that of a particular moral theory. Even if improving animal welfare were achievable (according to the criteria set out above), some routes of moral change would likely be more achievable than others. Creating psychologically plausible theories can help identify which routes are likely to meet stronger resistance than others.

Third, setting aside the issue of guidance, psychological plausibility matters insofar as we're committed to a standard moral methodology like reflective equilibrium. According to that methodology, it's a virtue of a theory that it captures our considered moral judgments. A theory might be more virtuous still insofar as it captures our considered moral judgments about high stakes practices—for instance, cases where humans or animals are vulnerable to particularly serious kinds of harms, where it's especially important that a candidate theory get things right. Psychologically plausible theories are likely to fare well in these respects; by design, they account for those features of human psychology that explain why certain moral judgments are fixed points for those who are doing the theorizing.

There is much more to say about, and in defense of, psychological plausibility, and some additional virtues of psychological plausibility should become apparent in our analysis of empirical research on attitudes toward animals. However, this is not a paper on psychological plausibility per se. So at this juncture, we simply want to register that we are taking for granted that psychological plausibility is a theoretical virtue that's relevant to moral theory selection. Our main aim, after all, is to see what follows from this assumption, and while we take it to be a plausible starting point, our goal here isn't to offer a systematic case for it.

4 Differential attitudes toward humans and animals

Let's grant, then, that it's a virtue of a moral theory that it's robustly psychologically plausible. Our next task is to argue that people make firm moral distinctions between humans and non-humans.

Research on human attitudes toward animals has generally found that we treat animals as categorically different from humans. Even at a phylogenetic level, a

classic line of research going back to Carey (1985) has argued that children require significant directed instruction to learn that humans are also animals. For example, 5-year-olds are willing to classify birds and dogs as similar to humans but 3-year-olds are not, suggesting that 3-year-olds have not yet developed a concept of shared animal nature between humans and non-humans (Herrmann et al., 2012). Leddon et al. (2012) conclude, based on a literature review, that across cultures the idea that humans are properly classified as animals doesn't develop until around the age of nine. In short, the distinction between humans and non-humans seems to be cross-cultural, present early in childhood, and persists unless children are formally taught to think otherwise.

Moreover, when we consider research that's supposed to show that children have *less* of a pro-human bias than adults, we find that it still establishes a notable pro-human bias. In a study of children ages 5–9, Wilks et al. (2021) presented children with a choice of saving various numbers of humans or various numbers of animals. Roughly 25% of children said that one human should be saved over 100 pigs, over 30% said that one human should be saved over 10 pigs, and over 50% said that one human should be saved over two pigs. Additionally, 10–15% of the children weren't sure what to say in each case. This suggests that even when the numbers are in animals' favor, many children are at least inclined to privilege humans. (Moreover, that lack of confidence may reflect a deeper problem in the study. If there is enough confusion about the question, which would not be surprising given the age of the respondents, then noise alone could produce the results that the experimenters found.)

By adulthood, this tendency is far more pronounced, and adults tend to judge animals, as a category, to be less morally significant than humans. For example, Cavioli et al. (2019), in a study of 140 US American participants, found that people were willing to relieve human suffering at roughly twice the rate that they would relieve the suffering of animals. They asked participants to donate $100 to either a charity that would help humans or one that would help animals. Donating $10, they were told, would keep one individual (human or animal) free from pain and suffering for a day. On average, the human-focused charity received $68 while the animal charity received $32. Donations were furthermore more highly correlated with beliefs about human superiority than beliefs about humans' greater intelligence or capacity to suffer. (Further reinforcing other parts of our argument here, this study also found much greater willingness to donate to mentally severely disabled humans than to chimpanzees, and to dogs over pigs.)

Beliefs about human superiority to other animals have been documented throughout history and across cultures, making it unlikely that such beliefs are highly contingent or easily modifiable features of contemporary societies.[3] Beliefs in human superiority may be so prominent because they stem from the ways we

---

[3] As Ingold (2000) and other anthropologists have argued, some cultures' perceptions of animals diverge from the majority views expressed in the kinds of psychological studies discussed here. Though beliefs about human superiority can be found across a wide variety of cultures, there is also significant diversity in those beliefs. The results discussed here should be understood as being particularly pronounced in industrialized cultures.

monitor membership in our ingroup. For example, research on dehumanization argues that we treat both certain humans and non-humans as inferior because we view them as a threat or in conflict with us (Haslam & Loughnan, 2014; Leyens et al., 2001). In the most extreme case, we dehumanize by asserting that someone lacks the core traits that make someone human. Even though animals are not humans, we can "dehumanize" them by suggesting they lack even the most basic emotions, like pain and pleasure. This style of dehumanization is especially common when justifying meat consumption by ascribing fewer mental states to animals (Bastian et al., 2012). For both animals and the humans we view as members of an outgroup, dehumanization helps justify poor treatment.

Similarly, research on social dominance orientation proposes that we view animals as inferior because we fundamentally view the world in hierarchical terms (Costello & Hudson, 2014). We have general-purpose mechanisms that seek to promote those we view as part of our ingroup and dominate those we see as members of an outgroup. Ingroup membership is typically determined by various behaviors and obvious physical features—for example, the language someone speaks or their skin color. We judge animals to be fundamentally different because they lack almost all of the normal indicators that something is "like us." As a result, people who are especially committed to a hierarchical worldview, and base their moral decisions on whether someone is part of their ingroup, are more likely to exclude nonhumans from moral concern (Waytz et al., 2019). They're more likely to think that animals can be used for human benefit, that animals are inherently inferior to humans, and to view improved treatment of animals as threatening (Dhont & Hodson, 2014).

Having said all this, we might still wonder whether the human-animal divide is an ineliminable feature of human psychology. Nothing we've said establishes that it is. Still, the above is good evidence that it's a stable and significant feature of human psychology *now*. And to achieve robust psychological plausibility, an ethical theory needs to address the obvious psychological constraint posed now and in any foreseeable future. This means taking into account, at least to some degree, differentiated attitudes toward humans and animals that track the species boundary.

This is where psychologically plausible hybrid theories come apart from existing approaches to animal ethics that appeal to the possession or absence of characteristics that are obviously morally relevant, such as agency, self-consciousness, etc. The familiar difficulty is that capacities like agency and self-consciousness do not track the species boundary. Babies, children, and people with severe mental disabilities may lack the relevant kind of agency; great apes, elephants, and some marine mammals may possess it. But a divide that includes some animals on the same side as most humans, while excluding other humans who lack the relevant capacity, lacks psychological plausibility, given the account above. The kind of hybrid theory we're thinking of uses human psychology as part of the explanation for why different normative principles apply to different classes of beings.

We're inclined to think, though, that hybrid theories don't go far enough in postulating just two moral domains—recalling that we've been focusing on theories that are structured around the human/animal divide. While this divide is important, there are others that deserve our attention. To bring out some of the relevant issues

here, we'll consider the limitations of Nozick's "Utilitarianism for animals, Kantianism for people." We suggest that a psychologically plausible hybrid theory needs to draw finer distinctions, and invoke more nuanced principles, than Nozick's does.

5 Some difficulties with "utilitarianism for animals, kantianism for people"

Recall Nozick's "Utilitarianism for animals, Kantianism for people," according to which we should "(1) maximize the total happiness of all living beings [and] (2) place stringent side constraints on what [we] may do to human beings." In other words:

> Human beings may not be used or sacrificed for the benefit of others; animals may be used or sacrificed for the benefit of other people or animals only if those benefits are greater than the loss inflicted… One may proceed only if the total utilitarian benefit is greater than the utilitarian loss inflicted on the animals. This utilitarian view counts animals as much as normal utilitarianism does persons. Following Orwell, we might summarize this view as: *all animals are equal but some are more equal than others.* (None may be sacrificed except for a greater total benefit; but persons may not be sacrificed at all, or only under far more stringent conditions, and never for the benefit of nonhuman animals…)

Let's focus on the "utilitarianism for animals" part of this view, though what we say will have implications for the "Kantianism for people" part as well. There are three reasons to think that "utilitarianism for animals" fits poorly with common judgments. First, (a) while most forms of utilitarianism are fully egalitarian, maintaining that similar interests should be taken into account equally across all species and contexts, people have strong anti-egalitarian intuitions about animals of different kinds and in different contexts. Second, (b) utilitarianism is committed to welfarism (primarily understood either in terms of hedonic states or preference satisfaction), which many people reject, in multiple ways, with respect to at least some animals. Third, (c) the point of "utilitarianism for animals" is to reject side constraints for animals, but that idea fits very poorly with most people's judgment about some animals, namely *companion* animals—the animals with whom we live in our homes.

5.1 Anti-egalitarian intuitions

Utilitarianism has one key egalitarian feature: it accepts the principle of equal consideration of interests, according to which similar interests deserve equal consideration, regardless of whose interests they are. However, studies show that people have strongly anti-egalitarian intuitions about animals—not just in relation to humans, but between different kinds of animals. Animals can be valued very differently based on a wide range of factors: their capacities, their aesthetic appeal,

the degree to which they benefit or threaten human beings, and so on. There's clear evidence, for instance, that people have strong psychological preferences for animals that are human-like in their physical appearance. We show preferential treatment toward animals with large eyes, colored irises, bony cheeks, and a "smile" resembling that of human beings (Herzog, 2010; Hecht & Horowitz, 2015; Horowitz & Bekoff, 2007). What's more, numerous willingness-to-pay surveys indicate that people would be willing to pay much more to protect members of some species than others, judgments that appear to be based on factors such as cultural value, cuteness, charisma, etc. (e.g., Wallmo & Lew, 2012).

Another example is what's sometimes called an "anthropogenic Allee" effect, where rare species are valued more highly than common species (Angulo & Courchamp, 2009; Courchamp et al., 2006). This implies that the value of individual animals—and the concern due them—is inversely proportionate to global population size (brutally illustrated in the wildlife trade, where prices for animal parts often escalate as animals become rarer). This value preference for individuals that are rare over those that are common is frequently found in conservation practice. For example, in the 1990s, the Galapagos Conservancy had sharpshooters kill 250,000 goats across the Galapagos Islands to save 15,000 tortoises—the famous (or infamous) Project Isabella. Since mammals are generally viewed as more cognitively sophisticated than reptiles, it is unlikely that that the goats' interests were discounted due to relative judgments about cognitive capacity (Phillips & McCulloch, 2005; Phillips et al., 2012; Sevillano & Fiske, 2016). Rather, the value of the Galapagos tortoises was driven by their rarity. We can see similar convictions play out in any number of conservation examples, where there's willingness to kill plentiful and relatively cognitively sophisticated animals to save rare and relatively unsophisticated ones. Moreover, when this sort of reasoning is resisted, it usually isn't based on the thought that similar interests should be taken equally into account, whoever has them. Instead, it's based on broadly deontological reasoning: that conservation practices should rarely involve seriously harming or killing sentient animals (e.g., Wallach et al., 2015).

5.2 Anti-welfarist intuitions

Many utilitarians are committed to welfarism: the view that welfare is the only thing with intrinsic moral value, so that all other values must be explained in terms of their impact on welfare. But there is ample evidence that wild animals suffer a great deal in the wild due to hunger, thirst, disease, injury, environmental conditions, predation, and any number of other factors. However, all widely accepted forms of wild animal management either allow or require non-intervention in cases where this suffering takes place. What's more, some forms of management require increasing the threat of suffering, such as predator reintroductions. This suggests widespread strongly anti-welfarist intuitions about wild animals, or at least the view that animals' wildness or ability to perform natural behavior, and the functioning of wild systems, take priority over the welfare of individual animals. We can make the same point from a different angle when we consider many arguments against keeping animals in zoos, which tend not to be based on worries that zoo animals are

worse off than they would be in the wild, but simply because life in captivity is thought to be inappropriate for wild animals (Gruen, 2014). Trying to pull *wild* animals into a hedonistic utilitarian framework would seem to require significant psychological shifts, and it isn't clear what psychological mechanism could get us there.

5.3 Companion animals: side constraints and special privileges

Finally, utilitarianism rejects side constraints—that is, absolute restrictions on what can be done—in principle (though it's compatible with having them in a decision procedure). Of course, numerous studies have shown that people view friends and relatives as having greater moral significance than strangers and non-kin (Holyoak & Powell, 2016). For example, when faced with a moral dilemma where they are asked to sacrifice one to save five (utilitarian reasoning), people are less likely to do so if the person to be sacrificed is someone close to them (Bleske-Rechek et al., 2010; Swann et al., 2010). Utilitarians can accommodate this to some degree, but they don't want to accommodate it to the degree that would be required to match ordinary intuitions, as evinced by standard utilitarian arguments for sacrificing for distant strangers as opposed to attending to the interests of those near and dear (a reason why utilitarianism is argued to be psychologically implausible for humans, too). This creates a problem when it comes to companion animals. As Kasperbauer (2017a, 43) notes, companion animals are often treated as "honorary members of human ingroups." As a result of this, people have intuitions about companion animals that seem close to being Kantian, or are in any case certainly not straightforwardly utilitarian. Many studies indicate that people regard their animal companions as family members (e.g., Power, 2008). For instance, 13 of the 16 pet owners interviewed in Cohen (2002) said they would give a scarce drug to their pet in preference to a person outside their community.

This suggests that companion animals benefit both from side constraints and special obligations; and there's plenty of evidence of this privileging in practice. For instance, pet spending in the U.S. in 2017 was estimated to exceed $72 billion, more than the combined GDP of the 39 poorest countries in the world.[4] While these expenses no doubt improve the welfare of the companion animals concerned, not to mention the happiness of human family members, they don't seem like a good fit with "Utilitarianism for animals, Kantianism for people." If there were more general utilitarian intuitions about animal companions, then we would expect very different behavior. Instead of being willing to pay a great deal to have a hip surgery performed on an elderly Labrador, people would peacefully euthanize the Labrador, using the money instead to assist dozens of other animals—say street dogs—whose welfare could be massively improved by redirecting those funds. But of course, very few people would be inclined to think that it's wrong to perform the surgery *for this reason*. If they think the surgery is objectionable, it's going to be because they

---

[4] See: https://www.americanpetproducts.org/press_industrytrends.asp and https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?year_high_desc=true.

worry about the welfare of the elderly Labrador, or because they think the money should be spent on people—not because they are utilitarians about animals. Our treatment of companion animals suggests widespread acceptance of a permission *not* to maximize utility.

So, while Nozick's hybrid does pick up on some psychological constraints around human value and treatment of animals, it's nonetheless insensitive to some deeply engrained features of human psychology. What characteristics, then, would a "better hybrid theory" have?

6 Towards a better hybrid

Hybrid theories can accommodate the kind of systematic patterns in human psychology we've identified here. Nozick's hybrid tries to accommodate a particularly strong feature of our moral psychologies: namely, how we treat humans compared to nonhumans generally. But Nozick's hybrid still fails to account for significant differences in how we value and treat particular humans and certain nonhumans, most notably those with whom we have special relationships. So, the goal of this last section of the paper is to think about what, in the context of human-animal relationships, the underpinnings of a better psychologically plausible hybrid theory might be.

As we mentioned earlier, though, it is important not to be overly concessive to the status quo. The general goal is to balance psychological plausibility with aspiration, to find the mean between adoptability and moral ambition.[5] This means, in essence, that the hybrid theory should be as demanding as psychological plausibility allows. More fundamentally, the theory should be as demanding as psychological plausibility allows when it comes to the consistency that it requires of us; it should force us to be more principled than we currently are.[6] It's also worth recognizing that psychological plausibility is, at least to some degree, relative to the amount that we are, collectively, willing to invest in psychological, sociological, and institutional change. Moral change is likely to be easier through collective rather than individual action.

With this in mind, what might underpin a viable proposal for a hybrid theory for animal ethics? Again, we are more interested in providing a proof of concept—

---

[5] We won't say much here about the grounds for moral ambition, other than that, in the cases we're concerned about, it shouldn't be terribly controversial. Practically all moral theories, including various contemporary forms of Kantianism and some kinds of contractarianism, agree that certain ways of treating animals are morally unacceptable; there is, for instance, no real debate about the moral merits of many of the practices that constitute intensive animal agriculture. Where we have that kind of agreement, we can plausibly say that the aim of moral ambition is theoretically neutral.

[6] It's probably true, though, that people value consistency with respect to human beings much more than they value consistency with respect to animals. So, egalitarian views may be more psychologically plausible for humans than across species. Since we are able to stay more motivated to invest in change that makes the world more egalitarian for human beings, we can have more ambitious ideals for humans; since we are unable to stay as motivated to invest in in change that makes the world more egalitarian for animals, we can't have as ambitious ideals.

evidence that it's possible to construct such theories—than to defend a particular proposal. So, what we sketch here is just a tentative hypothesis. Our goal is to show that there are combinations of domains and principles that are both reasonably psychologically plausible and involve some moral ambition. Ultimately, of course, it would take much more empirical work to establish that we've identified the relevant domains and the right principles. Still, our suggestions, based on current empirical evidence, is that there are at least three broad domains in our moral psychologies for which we must account if we want to create a psychologically plausible hybrid theory. They include (1) either human or honorary humans; (2) animals we use for various purposes, specifically including agriculture and research; and (3) wild animals. That is, we think that a viable proposal probably won't simply be a hybrid—if understood as involving two moral domains—but at least a "tribrid," recognizing three. What sorts of principles might govern these domains?

6.1 Humans and companion animals

We already noted that for many people, companion animals are part of their "ingroups," and they readily make the kinds of sacrifices for them that people regularly make for their children. Indeed, we've extended our norms to protect not just the mental and physical well-being of companion animals but even their dead bodies. In the U.S., almost everyone would be revolted by the thought of eating their dead dog or cat, though they may have no problem at all eating a dead pig or cow. Companion animal corpses, like human corpses, simply aren't things to be eaten. We submit that the existence of such taboos is good evidence that some sort of broadly deontological approach is the default setting for human relationships with companion animals. It's psychologically implausible to ask people to regard their "ingroup" companion animals in, for instance, a direct welfarist utilitarian framework. "Utilitarianism for animals" excludes too many animals to whom we are strongly attached and whom we want to treat as special.

Someone might argue that the moral significance of ingroup membership will diminish over time, such that future human beings will value only a being's welfare, not its closeness to us. However, ingroup biases seem so fundamental in human psychology that any theory proposing they could be altered would have to explain in some detail why we should think that's a real possibility. Such explanations haven't been forthcoming. So as things stand with respect to companion animals, we think some kind of deontological framework, of a similar kind to the one in which we include human beings, is likely to be the most psychologically plausible theoretical framework.

Someone might object that we are neglecting the very constraint for which we've argued: namely, the species boundary. By saying that some animals are honorary humans, aren't we denying that there really is a sharp human/animal divide? No. The crucial point here is that for animals to get this special treatment, they have to be counted as honorary humans; they have to be incorporated into human lives in some very significant way. When it comes to companion animals, their moral significance isn't due to an appreciation for their capacity to suffer, or some other feature that they share with the animals in farms and forests. Instead, it's due to

people having established deep emotional relationships with them, forming bonds of trust and shared expectations. Clearly, partiality and intimacy are doing the work here, and those things can't be scaled to all animals generally. So there really is a sharp human/animal divide, and insofar as it can be bridged, it's only by some of the most powerful forces that bind us to others.

That being said, we grant that the status of companion animals is contentious in society. Some people have incredibly deep attachments to their companion animals, treating them with many of the same privileges that they give to human family members, and there are plenty of other people who are willing to respect these privileges as legitimate. At the same time, there are lots of people who think of it as a deep mistake to give companion animals such a status, insisting that nonhuman animals should never be elevated to the same plane as human beings. It seems to us that, on balance, society is trending toward the former view, but that's an empirical hypothesis that could be mistaken. And if it is, we will need to split this domain into two: one for human beings, the other for companion animals. As we said earlier, though, the goal here is not to establish that this particular tribrid is the correct hybrid view. Instead, we are simply trying to argue that hybrid views have a better claim to be psychologically plausible than uniform ones, which requires outlining a position that can make that judgment plausible.

6.2 Animals we use in laboratories, agriculture, and related contexts

Attitudes towards the animals that we use, for instance for food and experiments, are, in general, very different from the attitudes we have towards the animals we keep as companions. Evidence suggests that people are much more likely to regard animals in the contexts of farms and labs in a utilitarian way, as Killoren and Streiffer (2020) note. Indeed, research animals seem to be the paradigm case here, as animal research is already frequently governed by frameworks that weigh the suffering of research animals alongside the benefits to humans (and by not exposing humans to similarly invasive research, implicitly place stringent side constraints on what may be done to humans). There's evidence from public surveys (e.g., Ipsos MORI 2018) that the majority of ordinary people already do take a broadly utilitarian attitude to animal research: people think that animal welfare should be protected as far as possible, the serious human benefits of medical research should be the primary justification, and the same goals should not be achievable by alternative means.

Changing this broadly utilitarian approach for laboratory animals—by moving to, say, some kind of deontological view with side-constraints—may seem relatively achievable. After all, there have been species who have gained protections over the years, such as chimpanzees and other primates. We're less optimistic, however. First, primate research is arguably the most controversial research done on animals, and yet after it became clear that monkeys were ideal models for COVID-19 vaccine development, they were in such high demand that there were monkey shortages around the world.[7] This suggests that progress for animals in these

---

[7] https://www.theatlantic.com/science/archive/2020/08/america-facing-monkey-shortage/615799/.

contexts is fragile—it can be undermined when the stakes are high. Second, it's important to recall that the "side constraints" for research animals are hardly equivalent to the side constraints we apply to human beings. Researchers can't cause *unnecessary* pain, but they can cause *a lot* of pain as long as they are studying something that requires it. So, for instance, if researchers are studying the impact of arthritis on mobility, they can easily get approval to do that study without a requirement to provide pain relief at all times. The painfulness of the arthritis is one factor that affects mobility and eliminating it would complicate the interpretation of the results. So, we're inclined to think that we are a long way from abandoning a broadly utilitarian approach to these animals, and that it's psychologically implausible to require it for that reason. But what kinds of ethical underpinning would both preserve plausibility while expressing some moral ambition, of improving things for these animals?

First, following Killoren and Streiffer (p. 6), who suggest the view but don't endorse it, we propose as one possible underpinning principle: Optimific Is Always Okay:

> Optimific Is Always Okay: It is never wrong, and hence it is always permissible, for an agent to act optimifically (where an action is optimific iff its consequences are at least as good as the consequences of any alternative action open to the agent).

Optimific Is Always Okay captures the idea that consequentialist reasoning is appropriate for these kinds of animals, though maximizing (whether well-being or something else) isn't required. However, all on its own, Optimific Is Always Okay doesn't go far enough; it lacks necessary moral ambition, because it lacks a floor for animal welfare (It doesn't say, for instance, that optimific is always *required*.) So, some kind of baseline is also needed, perhaps along the lines of Good Enough Is Always Required:

> Good Enough Is Always Required: If animals intended for human use are under your care and you have control over the factors that affect their welfare, then it is wrong for you to allow their welfare to fall below some threshold.

Obviously, Good Enough Is Always Required doesn't tell us what the crucial threshold is, but that's a feature rather than a bug. We suspect the standard is a moving target relative to the current state of knowledge regarding animal welfare and the kind of human/animal divide that's psychologically entrenched. This might be a problem if we weren't committed to psychological plausibility, as we might want a more principled answer based on the facts about what makes animals' lives of adequate quality. However, since psychological plausibility involves some kind of publicity condition—people have to be able to know what they're obligated to do—it's perfectly appropriate to include epistemic relativity of this kind.

6.3 Wild animals

As we've already noted, the situation with wild animals is different again; these animals are generally considered primarily as ecosystem or species members, even

if they are sentient. The belief that nature has intrinsic (non-instrumental) value, and wild animals have value as part of that nature, appears to be widespread (de Groot, Drenthen, & de Groot, 2011): people feel that nature is sacred and pure in a way that requires protection (Rottman, 2014; Rottman, Kelemen, & Young, 2015). When asked to donate money to protect various animal species, people tend to give more for species that are rare, endangered, or play a crucial role in preserving ecosystems (Lew, 2015; Wainger et al., 2018). So, the importance of these wild animals is measured in terms of their contribution (whether positive or negative) to valued ecosystems, or their role as supporting or undermining valued species. This may also mean that where wild animals are found in human dominated systems, whether agricultural or urban, they are frequently seen as pests.

Generally, people have been willing to allow individual wild animals to be harmed and killed to promote both human and conservation goals (such as promoting crop or species protection or ecosystem health). Such an ethical approach to wild animals is far removed either from the deontological constraints for companion animals or the two principles we suggested for research and agricultural animals: Optimific Is Always Okay and Good Enough is Always Required.

Some kind of theoretical approach that prioritizes ecosystem and species protection and promotion, or human wellbeing (in the case of homes and crops) would best reflect what's psychologically plausible here. However, we still think there's room for moral ambition. One way of reflecting this would be to defend a theoretical approach that requires humaneness in the ways in which these goals are pursued. At present this is not always the case (for instance, in using poisons such as 1080 to kill invasive species for conservation purposes). An underpinning principle that reflects both psychological plausibility and moral ambition might be If You Intervene Ecologically, Intervene Humanely.

> If You Intervene Ecologically, Intervene Humanely: It is only permissible to optimize ecosystem health if it's done as humanely as possible.

This principle imposes a side constraint on environmental action. It grants that it's okay to do what's best for ecosystems even if that involves harms to nonhuman animals; the only requirement is that the intervention doesn't cause more suffering than necessary. Plainly, that may involve causing some suffering, and it also involves tolerating all the pain inherent in nature. Still, it doesn't challenge the widespread idea that we don't act wrongly when we do what's good for ecological wholes even if that means setting back the interests of some animals.

That said, based on public responses to a wide variety of anthropogenic or partly anthropogenic disasters, we think that it's psychologically plausible to be a bit more morally ambitious with wild animals, at least in the case of large-scale impacts on ecosystems. Where humans have caused large-scale harms to wild animals—for instance due to oil spills—there's generally support for assisting them to recover from these harms. We think this can be systematized into another underpinning for wild animals: Make Good Your Harms.

Make Good Your Harms: If you are responsible for large-scale harms to wild animals in wild ecosystems, then you have some duty of repair to those wild animals.

Make Good your Harms fits neatly with the idea that, if there's a significant pollution incident involving an oil spill, the company ought to pay to save or humanely euthanize, if necessary, the seals who get covered in it. Moreover, it captures the idea that the obligation isn't simply based on the importance of expressing a certain virtue; you actually owe something to the animals themselves. Make Good Your Harms centers animals by recognizing them as individuals to whom things can be owed, even in a context where their primary value is to the ecosystem or species. This may be an area, in fact, where at least in post-industrial countries, public attitudes are shifting; in Kasperbauer's terms, psychological profiles seem to be transitioning. In recent research, Manfredo et al. (2020) for instance, found a shift towards "mutualistic" values for wildlife where wild animals are increasingly seen as "part of one's social network and worthy of care and compassion." This transition would certainly make principles like Make Good Your Harms seem increasingly psychologically plausible, and would also make for more acceptance of humane responses to "pests" and other out-of-place wildlife.

7 Conclusion

There are, no doubt, philosophers who deny that psychological plausibility is a theoretical virtue. Many animal ethicists seem to be among them, insisting that no matter how counterintuitive it might be, we should give equal consideration to the like interests of all animals, human and nonhuman, whether they are in homes, zoos, farms, laboratories, or wild forests. Indeed, this argument seems to lie behind the growth in claims that we should act to relieve wild animal suffering from causes such as disease and predation (e.g., Horta, 2016). But insofar as we regard psychological plausibility as a theoretical virtue, we have some reason to favor more modest moral theories. As things currently stand, human psychology seems to push us toward something like a "tribrid" view, placing humans and honorary humans in one camp, research and agricultural animals in another, and wild animals in a third.

Of course, the principles that we've suggested for these domains only provide an outline of what such a tribrid theory might look like. We've glossed over many issues that deserve greater attention, and we grant, in particular, that there may be more than three domains into which we divide animals (perhaps splitting humans and companion animals into two domains, for instance, or adding in a special domain for the "liminal" animals who live around our homes and in our cities— those to whom Donaldson and Kymlicka [2011] refer as "denizens"). Our goal here has simply been to show that hybrid theories can be psychologically plausible, and we hope that our tribrid could be a starting point for further work in this area.

References

Angulo, E., & Courchamp, F. (2009). Rare species are valued big time. *PLoS ONE, 4*(4), e5215. https://doi.org/10.1371/journal.pone.0005215

Bastian, B., Loughnan, S., Haslam, N., & Radke, H. (2012). Don't mind meat? The denial of mind to animals used for human consumption. *Personality and Social Psychology Bulletin, 38*, 247–256.

Bleske-Rechek, A., Nelson, L. A., Baker, J. P., & Brandt, S. J. (2010). Evolution and the trolley problem: People save five over one unless the one is young, genetically related, or a romantic partner. *Journal of Social, Evolutionary, and Cultural Psychology, 4*, 115–127.

Carey, S. (1985). *Conceptual change in childhood.* MIT Press.

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, 116*, 1011–1029.

Cohen, S. (2002). Can pets function as family members? *Western Journal of Nursing, 24*(6), 621–638.

Costello, K., & Hodson, G. (2014). Lay beliefs about the causes of and solutions to dehumanization and prejudice: Do nonexperts recognize the role of human-animal relations? *Journal of Applied Social Psychology, 44*, 278–288.

Courchamp, F., Angulo, E., Rivalan, P., et al. (2006). Rarity value and species extinction: The anthropogenic Allee effect. *PLoS Biology, 4*(12), e415. https://doi.org/10.1371/journal.pbio.0040415

De Groot, M., Drenthen, M., & de Groot, W. T. (2011). Public visions of the human/nature relationship and their implications for environmental ethics. *Environmental Ethics, 33*, 25–44.

Dhont, K., & Hodson, G. (2014). Why do right-wing adherents engage in more animal exploitation and meat consumption? *Personality and Individual Differences, 64*, 12–17.

Donaldson, S., & Kymlicka, W. (2011). *Zoopolis: A political theory of animal rights.* Oxford University Press.

Flanagan, O. (1991). *The science of the mind.* Cambridge: MIT Press.

Gruen, L (Ed.). (2014). Dignity, captivity, and an ethics of sight. In L. Gruen (Ed.), *The ethics of captivity* (pp. 231–247). New York: Oxford University Press.

Hamlin, A., & Stemplowska, Z. (2012). Theory, ideal theory and the theory of ideals. *Political Studies Review, 10*, 48–62.

Haslam, N., & Loughnan, S. (2014). Dehumanization and infrahumanization. *Annual Review of Psychology, 65*, 399–423.

Hecht, J., & Horowitz, A. (2015). Seeing dogs: Human preferences for dog physical attributes. *Anthrozoös, 28*(1), 153–163.

Herrmann, P., Medin, D. L., & Waxman, S. R. (2012). When humans become animals: Development of the animal category in early childhood. *Cognition, 122*, 74–79.

Herzog, H. (2010). *Some we love, some we hate, some we eat.* New York: Harper Perennial.

Holyoak, K. J., & Powell, D. (2016). Deontological coherence: A framework for commonsense moral reasoning. *Psychological Bulletin, 142*, 1179–1203.

Horowitz, A. C., & Bekoff, M. (2007). Naturalizing anthropomorphism: Behavioral prompts to our humanizing of animals. *Anthrozoös, 20*(1), 23–35.

Horta, O. (2016). Animal suffering in nature: The case for intervention. *Environmental Ethics, 39*(3), 261–279.

Ingold, T. (2000). *The perception of the environment.* Routledge.

Ipsos MORI (2018). *Public attitudes to animal research.* Available at https://www.ipsos.com/sites/default/files/ct/news/documents/2019-05/18-040753-01_ols_public_attitudes_to_animal_research_report_v3_191118_public.pdf

Kasperbauer, T. J. (2017a). Mentalizing animals: Implications for moral psychology and animal ethics. *Philosophical Studies, 174*, 465–484.

Kasperbauer, T. J. (2017b). *Subhuman: The moral psychology of human attitudes to animals.* New York: Oxford University Press.

Killoren, D., & Streiffer, R. (2020). Utilitarianism about animals and the moral significance of use. *Philosophical Studies, 177*(4), 1043–1063.

Leddon, E., Waxman, S.R., Medin, D.L, Bang, M., & Washinawatok, K. (2012). One animal among many? Children's understanding of the relation between humans and nonhuman animals. In G. Hayes & M. Bryant (Eds.), *Psychology of culture* (pp. 105–126). Hauppauge, NY: Nova Science Publishers.

Leyens, J.-P., Rodriguez-Torres, R., Rodriguez-Perez, A., Gaunt, R., Paladino, M., Vaes, J., & Demoulin, S. (2001). Psychological essentialism and the differential attribution of uniquely human emotions to ingroups and outgroups. *European Journal of Social Psychology, 81*, 395–411.

Lew, D. K. (2015). Willingness to pay for threatened and endangered marine species: A review of the literature and prospects for policy use. *Frontiers in Marine Science*. https://doi.org/10.3389/fmars.2015.00096

Manfredo, M. J., et al. (2020). The changing sociocultural context of wildlife conservation. *Conservation Biology, 34*(6), 1549–1559.

Nozick, R. (1974). *Anarchy, state and utopia.* New York: Basic Books.

Phillips, C. J. C., & McCulloch, S. (2005). Student attitudes on animal sentience and use of animals in society. *Journal of Biological Education, 40*, 17–24.

Phillips, C. J. C., Izmirli, S., Aldavood, S. J., Alonso, M., Choe, B. I., Hanlon, A., Handziska, A., Illmann, G., Keeling, L., Kennedy, M., Lee, G. H., Lund, V., Mejdell, C., Pelagic, V. R., & Rehn, T. (2012). Students' attitudes to animal welfare and rights in Europe and Asia. *Animal Welfare, 21*, 87–100.

Power, E. (2008). Furry families: Making a human-dog family through home. *Social and Cultural Geography, 9*(5), 535–555.

Rottman, J. (2014). Breaking down biocentrism: Two distinct forms of moral concern for nature. *Frontiers in Psychology, 5*, 905.

Rottman, J., Kelemen, D., & Young, L. (2015). Hindering harm and preserving purity: How can moral psychology save the planet? *Philosophy Compass, 10*, 134–144.

Sevillano, V., & Fiske, S. T. (2016). Warmth and competence in animals. *Journal of Applied Social Psychology, 46*, 276–293.

Stone, C. (1988). *Earth and other ethics*. Oxford University Press.

Swann, W. B., Gómez, A., Dovidio, J. F., Hart, S., & Jetten, J. (2010). Dying and killing for one's group: Identity fusion moderates responses to intergroup versions of the trolley problem. *Psychological Science, 21*, 1176–1183.

Valentini, L. (2012). Ideal vs. non-ideal theory: A conceptual map. *Philosophy Compass, 7*, 654–664.

Wainger, L. A., Helcoski, R., Farge, K. W., Espinola, B. A., & Green, G. T. (2018). Evidence of a shared value for nature. *Ecological Economics, 154*, 107–116.

Wallach, A. D., Bekoff, M., Nelson, M. P., & Ramp, D. (2015). Promoting predators and compassionate conservation. *Conservation Biology, 29*, 1481–1484.

Waytz, A., Iyer, R., Young, L., et al. (2019). Ideological differences in the expanse of the moral circle. *Nature Communications, 10*, 4389.

Wallmo, K., & Lew, D. K. (2012). Public willingness to Pay for recovering and downlisting threatened and endangered marine species. *Conservation Biology, 26*(5), 830–839.

Wiens, D. (2013). Demands of justice, feasible alternatives, and the need for causal analysis. *Ethical Theory & Moral Practice, 16*, 325–338.

Wilks, M., Caviola, L., & Kahane, G. (2021). Children prioritize humans over animals less than adults do. *Psychological Science, 32*(1), 27–38.