



Three and a half ways to a hybrid view in animal ethics

David Killoren¹ · Robert Streiffer²

Accepted: 8 March 2022 / Published online: 19 April 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract

The distinctive feature of a hybrid view (such as Nozick’s “utilitarianism for animals, Kantianism for people”) is that it divides moral patients into two classes: call them *dersons* and *uersons*. *Dersons* have a deontological kind of moral status: they have moral rights against certain kinds of optimific harms. *Uersons*, by contrast, have a utilitarian kind of moral status: their interests are morally important (in proportion to the magnitude of those interests), but *uersons* do not have deontological moral rights or any other kinds of deontological protections. In this paper, we discuss and critically evaluate three ways of supporting a hybrid view: a case-based argument; an autonomy-based rationale; and a rationale based in a capacity for what we call deep commitments. Finally, we discuss a way in which considerations about the moral significance of relationships might support an approximation of a hybrid view.

Keywords Animal ethics · Autonomy · Consent · Relationships

1 Introduction

In *Anarchy, State, and Utopia*, Robert Nozick describes a view that he dubs “utilitarianism for animals, Kantianism for people”:

It says: (1) maximize the total happiness of all living beings; (2) place stringent side constraints on what one may do to human beings. Human beings may not be used or sacrificed for the benefit of others; animals may be used or sacrificed for

✉ David Killoren
david.j.killoren@gmail.com

¹ Koç University, Istanbul, Turkey

² University of Wisconsin-Madison, Madison, WI, United States

the benefit of other people or animals *only if* those benefits are greater than the loss inflicted. (1974, 39)

Inspired by Nozick, Shelly Kagan considers a view he calls “restricted deontology,” according to which:

[a]n adequate moral theory must contain deontological elements (rights, for example), but the protections afforded by these deontological elements are not to be accorded to animals, only to people. ... Animals do count—their interests must be given weight—but for all that, when it comes to animals a different moral framework is called for. (2019, 191).

Both views are examples of Nozickian hybrid views—hereafter ‘hybrid views’ for short.¹ Hybrid views divide moral patients into two different classes. Call them *dersons* and *uersons*. Dersons have a deontological kind of moral status, so deontological principles determine how those individuals ought to be treated. Dersons have moral rights, which means that it is not always acceptable to harm them for the greater good; they enjoy some moral protections against optimific harm.² Uersons have a utilitarian kind of moral status, so utilitarian principles determine how those individuals ought to be treated. Uersons’ interests are morally important (in proportion to their magnitude), but uersons do not have deontological moral rights or any other deontological protections.

A full-fledged hybrid view, on our preferred conception, says not only that (i) some individuals ought to be treated according to deontological principles and that other individuals ought to be treated according to utilitarian principles, but also that (ii) this is so because of differences in moral status. A view that endorses (i) and denies (ii) is merely an ersatz hybrid view. Thus, for example, if a two-level utilitarian (e.g., (Varner 2012)) were to maintain that (iii) all beings who have interests have the same moral status, but that (iv) we ought to treat humans according to deontological principles while we ought to treat animals according to utilitarian principles, and that (v) we ought to do so because doing so will maximize utility, then they would have advanced an ersatz hybrid view, not a genuine one.

Different hybrid views will distinguish dersons from uersons on the basis of different properties. Some will draw the distinction in terms of membership in the species *homo sapiens*, such that all and only humans are dersons, and all and only animals are uersons.³ Others will draw the distinction in terms of autonomy or other psycho-

¹ Not all hybrid views need be ‘Nozickian’: one could take a contractualist approach to humans, a virtue-theoretic approach to non-human animals, and an ecocentric approach to plants, for example. Since our focus is on Nozickian hybrid views, we’ll use ‘hybrid view’ to refer to Nozickian hybrid views only.

² An act is optimific if and only if its consequences are at least as good as the consequences of any alternative action open to the agent.

³ Recall one of the oldest objections to utilitarianism, that utilitarianism is “a doctrine worthy only of swine” (Mill 1871, 121–22). This fairly straightforwardly implies that utilitarianism is in fact worthy of swine, even though it is not worthy of humans. On one plausible interpretation, this is an endorsement of a hybrid view.

logical characteristics, which may allow some animals to count as dersons or some humans to count as uersons.⁴

Further, different hybrid views will involve different principles governing our treatment of uersons and dersons.⁵ For instance, hybrid views may disagree about whether we ought to promote satisfaction of uersons' preferences or instead promote uersons' achievement of positive hedonic states. Similarly, hybrid views may disagree about what rights dersons have or about the circumstances, if any, in which those rights can be permissibly infringed.

These points illustrate that there is a vast range of possible hybrid views. We'll use *the Hybrid Theory* as our name for the idea that some hybrid view or other is correct.

Nozick and Kagan both reject the hybrid views they consider. In doing so, they are not unusual; hybrid views have few defenders in contemporary philosophical literature.⁶

However, the social norms, laws, and regulations in many institutions where animals are used seem to reflect a predominantly utilitarian approach, whereas utilitarian approaches in analogous human institutions are far less evident.⁷ For example, although many find utilitarian defenses of some animal agriculture compelling, almost everyone would deny that humans can be bred, raised, and slaughtered for food, even in hypothetical cases where the benefits of doing so outweigh the costs. Similarly, although utilitarian approaches are typical in mainstream ethical views of animal research, regulations governing human subjects research take a more deontological shape (Killoren and Streiffer 2020). Because some sort of a hybrid view seems to underlie many common practices, hybrid views continue to merit critical attention.

Developing a hybrid view that is both detailed and coherent is more complex and difficult than it may first appear (Kagan 2019, 191–214; Killoren and Streiffer 2020). But our task in this paper is to address a more fundamental issue: What can be said in favor of the Hybrid Theory? That is, what reasons are there for dividing moral patients into uersons and dersons in the first place? If no compelling answers can be found, then the Hybrid Theory and all particular hybrid views should be abandoned.

⁴ In *The Case for Animal Rights*, for example, Tom Regan allows that, while all normal mammals over 1 year of age are subjects-of-a-life and so have a basic right to respectful treatment, utilitarianism might be appropriate for sentient animals that are not subjects-of-a-life (1983, 246).

⁵ Utilitarianism and deontology themselves are large families of theories: just consider, e.g., satisficing vs. maximizing utilitarianism and egalitarian vs. prioritarian deontological theories.

⁶ The few include Loren Lomasky: "Human beings are *rights-holders*, and nonhuman animals are not. People may not be sacrificed to the greater good of others, while animals may" (2013, 195 emphasis in the original). Carl Cohen holds that humans have rights, that animals do not have rights, and that we still have direct duties to animals, but he fails to specify the principles governing our relations to animals in sufficient detail to know whether he is a utilitarian about animals (1986, 867; 1997, 95; 2001, 5–6, 38, 46). Similarly with Dario Ringach (2011, 309–10). Also, as noted in fn. 4, Regan suggests a hybrid view, albeit one according to which the class of dersons is significantly expanded and the class of uersons is significantly restricted.

⁷ This isn't to say that utilitarian approaches are completely absent in our dealings with our fellow human beings. For example, during the current pandemic, many countries restricted individual liberties in order to promote public health—a decision that is at least consistent with a utilitarian approach. Thanks to a reviewer for pushing us on this point.

We argue that there are several ways of trying to support the Hybrid Theory that need to be considered before philosophers can reach a final verdict about the prospects, or lack thereof, of the Hybrid Theory. Although this paper is organized around arguments *for* the Hybrid Theory, throughout the paper we highlight the weaknesses of these arguments. Our hope is to inspire further discussion and debate, rather than to settle the question of whether the Hybrid Theory is defensible.

In Sect. 2, we consider whether intuitions about cases support the Hybrid Theory, focusing on psychological research conducted by Lucius Caviola and colleagues. Caviola and colleagues believe that their research suggests that the folk do not embrace the Hybrid Theory. We argue that their results are inconclusive on this point; but we concede that any defender of the Hybrid Theory cannot appeal exclusively to intuitions about cases to support their view. In Sects. 3 and 4, we consider two novel rationales for the Hybrid Theory and explore challenges that they face. The first rationale appeals to the idea that the autonomy of some individuals but not others is intrinsically morally valuable. The second appeals to the idea that some individuals but not others have a capacity for what we call deep commitments. In the conclusion, we discuss how differences between human-animal relationships and human-human relationships relate to hybrid views. While those differences may support a view that resembles a hybrid view, it isn't clear whether that view should count as a hybrid view, strictly speaking.

2 Intuitions about Cases and the Hybrid Theory

Case-based intuitionism is the view that, *ceteris paribus*, theory T_1 is preferable to theory T_2 if T_1 is a better fit with our intuitions about cases than T_2 . Case-based intuitionism supports a methodology in which a theory's consistency with commonsense intuitions about cases confers a kind of presumptive authority. That is, a theory's consistency with intuitions is treated as a strong but not dispositive count in favor of it—a factor alongside other desirable features that theories can have (such as explanatory potency, consistency with widely accepted general principles, and simplicity). This methodology is widespread in contemporary ethics. Therefore, it's important to ask whether the Hybrid Theory is defensible on case-based intuitionist grounds. To explore this question, we'll consider empirical work conducted by Caviola, et al. (2021). We're focusing on their work because it represents the best currently available empirical evidence about whether ordinary people accept a hybrid view. We'll argue that their research suggests that a case-based intuitionist argument for *or against* the Hybrid Theory might be hard to produce.

To see what a case-based argument for (or against) the Hybrid Theory might look like, consider the following case template:

Outbreak Template: There's been an outbreak of a fatal virus. N_1 individuals of kind K are at risk of catching the virus and dying, and the only way to save their lives is to develop a vaccine. However, to develop the vaccine, N_2 healthy

individuals of kind K , who otherwise would survive, need to be infected with the virus, which will inevitably lead to their death.⁸

This template can be used to generate indefinitely many more specific cases. Intuitions about those specific cases might support the Hybrid Theory—or undermine it.

What would it take for intuitions about versions of Outbreak to support the Hybrid Theory? The following conditions would be sufficient. Firstly,

(i) for any Outbreak case involving some kind K_1 (e.g., animals) in which N_1 is larger than N_2 , we have the intuition that it is permissible to infect the N_2 to save the N_1 .

If (i), then our intuitions about instances of Outbreak would be in line with the view that individuals of kind K_1 are persons, and so treatment of individuals of kind K_1 ought to be guided by some utilitarian principle (in particular, some principle that entails that it is right to optimifically harm individuals of kind K_1 whenever only interests of individuals of kind K_1 are at stake).

Secondly, for intuitions about versions of Outbreak to support the Hybrid Theory, it seems that one of the following claims must be true: either

(ii) for any Outbreak case involving some kind K_2 (e.g., humans), we have the intuition that it is not permissible to infect the N_2 to save the N_1 (no matter how small N_2 is, and no matter how large N_1 is),

or

(iii) for Outbreak cases involving some kind K_2 (e.g., humans), we have the intuition that it is permissible to infect the N_2 to save the N_1 when, but only when, N_1 is *much* larger than N_2 .

If (ii), our intuitions about instances of Outbreak would support the view that individuals of kind K_2 are persons whose treatment is governed by an absolute deontological prohibition against causing the sort of harm involved in Outbreak.⁹ If (iii), our intuitions would support the view that individuals of kind K_2 are persons whose treatment is governed by some sort of “threshold deontology,” according to which certain kinds of harms to them are prohibited, even when optimific, unless such harms produce benefits exceeding some threshold.

Thus, if (i) and either (ii) or (iii) hold, then our intuitions would (from a case-based intuitionist perspective) provide support for the Hybrid Theory.

⁸ This is a version of a template used in experiments conducted by Caviola et al. (2021).

⁹ The qualifier “the sort of harm involved in an instance of Outbreak” is important, as some people (e.g., Thomson 1990, 292) hold that animals have deontological protections against being caused pain, but not against being killed. In this regard, one could have intuitions about animals that conform to (i) and yet still not be a utilitarian towards animals; one would only be a utilitarian about animal death.

Inconveniently for the defender of the Hybrid Theory, however, there seems to be evidence that our intuitions do *not* fit that neat pattern. Caviola et al. (2021) presented subjects with various versions of Outbreak involving humans or involving animals. Their results are consistent with the hypothesis that most people's intuitions fit the pattern in (ii) or (iii) with respect to humans but seem to undermine the hypothesis that most people's intuitions fit the pattern in (i) with respect to animals.

Kahane and Caviola explain some of their key findings:

[We] asked participants how many beings would need to be saved, at a minimum, in order to make it morally right to kill 10 beings of the same species [in a version of Outbreak]. Participants were also able to indicate that they consider it never right to make such sacrifices irrespective of the number of saved beings. We found that 65% of participants indicated that it is never right to kill 10 humans irrespective of the number of humans that could be saved. By contrast, only a minority of participants indicated that it is never right to kill 10 animals to save more animals of the same species. 34% indicated it was never right to kill 10 panda bears, 36% indicated it was never right to kill 10 dogs, 39% indicated it was never right to kill 10 squirrels, 30% indicated it was never right to kill 10 chimpanzees, and 25% indicated it was never right to kill 10 pigs. (n.d., 8–9)

So, when $N_2 = 10$, the intuitions of about two-thirds of participants in Caviola, et al.'s study support an absolute deontological constraint against (certain kinds of) harm to humans—consistent with (ii) above—whereas the intuitions of most participants do not suggest an analogous absolute deontological constraint against harm to any of the five non-human species mentioned in the experiment. So far, so good for the defender of the Hybrid Theory.

However:

Of those participants who indicated that there was a number of saved beings that would make it morally right to kill 10 beings of the same species, the mean responses (after adjusting extreme outliers using the winsorization technique) were the following: 201 humans, 64 panda bears, 60 dogs, 59 squirrels, 53 chimpanzees, and 51 pigs. (Kahane and Caviola, n.d., 9)

These results suggest that intuitions with respect to animals of most participants in Kahane and Caviola's study do not conform to (i). Participants' intuitions would most clearly conform to (i) only if they judged it to be morally right to kill 10 animals as long as this would save *11 or more* of the same kind of animal and that is not borne out in the study.

And so Kahane and Caviola argue that participants' intuitions do not fit with the Hybrid Theory and instead fit a version of threshold deontology that they call Multi-Level Weighted Deontology (MLWD), according to which:

there is a hierarchy of moral status and individuals that are lower in the hierarchy (e.g., pigs) can be sacrificed for the sake of those higher up (e.g., humans).

But within each level of moral status, the deontological constraints offer the same protections (i.e., it's generally wrong to sacrifice a pig to save five pigs), and these protections apply to the same degree. [Further, on this view,] deontological protections are not absolute, and get weaker the lower the level of moral status. (n.d., 7)

For our purposes, the important difference between MLWD and the Hybrid Theory is that MLWD implies that everyone, human or animal, is a derson. The Hybrid Theory, by contrast, implies that some individuals have qualitatively different moral statuses than others—some are dersons and others are ersons.

We agree that Kahane and Caviola's results cause problems for a case-based intuitionist defense of the Hybrid Theory and that those results suggest that a case-based intuitionist defense of MLWD may hold promise. But there are reasons for caution.

Consider what might be called *pseudo-deontology*. An approach is pseudo-deontological, let's say, when it yields verdicts that are characteristically deontological through reasoning that is either explicitly utilitarian or at least consistent with utilitarianism. The thought that utilitarian reasoning can support judgments of a deontological form has long been familiar in utilitarian moral philosophy.¹⁰

Pseudo-deontology is relevant here because if participants in Caviola et al.'s experiments take a pseudo-deontological approach to animals, then their intuitive judgments, despite being characteristically deontological, are consistent with their believing (or reasoning in accordance with) a hybrid view. To produce evidence bearing on whether participants' intuitions about animals are genuinely deontological or pseudo-deontological, it would be necessary to investigate the considerations that lead participants to their intuitions; intuitions on their own don't settle the matter. But that sort of investigation is not part of Caviola, et al.'s study (and is not typically part of psychological research of this kind).

Also, in interpreting their findings, Caviola et al. seem to rely on an excessively narrow understanding of what is involved in a hybrid view's approach to animals. They assume that such an approach implies that it is "completely permissible" and "absolutely morally right" to optimifically harm animals to save a larger number (2021, 1015). But the only commitment of hybrid views is that it is *permissible* to optimifically harm animals (Killoren and Streiffer 2020, 1048). One who accepts a hybrid view could think that the fact that an action causes pain to ten pigs, for example, is a strong reason not to do it, and thus not "completely permissible" or "absolutely morally right," yet still justified on balance by the lives saved as a consequence of the action. This is important because if one can deny that optimific harm is "absolutely morally right" without giving up a hybrid view, then the responses

¹⁰ As Mill notes, the view that there is one ultimate moral principle (e.g., the principle of utility) doesn't commit one to using that principle to "test each individual action directly." Instead, one can, and often should, use intermediate generalizations and secondary principles such as "One shouldn't murder" and "One shouldn't steal" (1871, 137–38). Sidgwick similarly argues at length for a utilitarian understanding of "the morality of common sense," despite its outwardly non-utilitarian appearance (1907, 423–59). Two-level utilitarianism is a development of this idea. For a recent articulation and defense of two-level utilitarianism, see (Varner 2012).

of participants in several of Caviola et al.'s studies might be more consistent with a hybrid view than Caviola et al. seem to allow.¹¹

Furthermore, the diversity of responses in Caviola et al.'s experiments muddies the waters. At least three separate theoretical approaches can be discerned in their results.

Caviola et al.'s data suggest that a sizable group of participants believe it is never right to kill the N_2 when $N_2=10$, no matter how large N_1 becomes—*regardless* of the species of animal: “34% indicated that it was never right to kill 10 panda bears, 36% indicated it was never right to kill 10 dogs,” and so on (Kahane and Caviola, n.d., 8–9). This suggests that a sizeable minority have intuitions that don't obviously conform to MLWD or the Hybrid Theory. Rather, the intuitions of individuals in this group seem to suggest some form of species-blind absolutist deontology.

Moreover, Kahane and Caviola report that 65% of their participants in one study and 83.8% in another indicated an absolutist approach to humans, that “it is never right to kill 10 humans irrespective of the number of humans that could be saved” (Kahane and Caviola, n.d., 8–9). Since MLWD, as defined by Kahane and Caviola, says that deontological constraints are not absolute, these responses do not conform to MLWD; but these responses *are* consistent with the Hybrid Theory.

Finally, their data suggest that there is a large, third group who say that, for all species, including humans, there is some number of saved individuals that can justify killing the N_2 in Outbreak. Most of these participants held that the number of saved individuals required to justify such killing is largest when the species in question is human and decreases for other species. Without further interrogation of these participants' reasoning, we cannot say whether their responses were given in a pseudo-deontological or genuinely deontological frame of mind; but it is fair to say that these participants gave responses that are at least consistent with the sort of deontological approach that falls out of MLWD.

The relative sizes of these three groups are unclear from the reports given by Caviola, et al., but it is at least clear that these groups are large and therefore that none of them can be claimed to represent the vast majority of respondents, as Kahane and Caviola agree (Kahane and Caviola, n.d., p. 13). For this reason, we think Caviola et al.'s results do not support the hypothesis that “the folk” accept or reject MLWD, the Hybrid Theory, or any other single view. Rather, their results suggest that ordinary people's intuitions about versions of Outbreak are diverse: some people's intuitions may be in line with MLWD, while other people's may be in line with the Hybrid Theory, and still other people's may be in line with neither.¹²

¹¹ In several of their studies, Caviola et al. use a 7-point Likert scale, which they describe as asking participants “to indicate how morally right (i.e., permissible) or wrong they thought it to be to harm 10 animals (humans) to save 100 animals (humans) on a 7-point scale from 1 (absolutely morally wrong) to 7 (absolutely morally right)” (2021, 1013). If a hybrid view is committed to saying that optimific harm is “absolutely morally right” then any response other than 7 is incompatible with a hybrid view; but if a hybrid view is only committed to saying that optimific harm is *on balance* morally right, then a larger range of responses would be consistent with a hybrid view. Another issue worth flagging here is that, if “right” is being interpreted as “permissible”, as the quoted language above says, then “(4) neither right nor wrong” means “(4) neither permissible nor wrong,” which is clearly not what is intended.

¹² There is one further concern that we wish to flag here. We think that case-based intuitionists should build their theories on the basis of *stable* intuitions—i.e., intuitions that persist after being informed of the

In this section, we have focused on Caviola et al.'s work because it currently represents the most direct empirical evidence bearing on whether ordinary people's intuitions fit with the Hybrid Theory. For the reasons we've given, we think that this evidence is ultimately inconclusive. So, we believe, if a good case for (or against) the Hybrid Theory is to be made, it probably can't rest on intuitions about cases like Outbreak. In the rest of this paper, we'll discuss other sorts of arguments for the Hybrid Theory.

3 Autonomy, Value, and Hybrid Views

The defender of the Hybrid Theory has two challenges: (i) identifying a property which differentiates persons from non-persons and (ii) explaining how this difference can explain why the beings they classify as persons (humans, typically) are in fact persons and why the beings they classify as non-persons (animals, typically) are in fact non-persons. Shelly Kagan's recent discussion of a hybrid view focuses on whether autonomy could serve as such a property. In this section, we'll first examine Kagan's discussion of autonomy and argue that his argument that animals have less autonomy than humans is unsound (3.1). Then we'll produce and evaluate a new rationale for the Hybrid Theory based not on a difference between humans' and animals' autonomy, but rather on a difference between the *value* of humans' and animals' autonomy. The first stage of that rationale involves an argument that animals' autonomy is not intrinsically valuable (3.2). The second stage argues that humans' autonomy *is* intrinsically valuable and that its value is not morally fungible (3.3).

3.1 Kagan on the Autonomy of Animals

Kagan explains what it means to be autonomous:

an autonomous being has preferences about how they want their life to go, preferences that are neither simply imposed by external forces or circumstances, nor merely a matter of instinct rather than individual choice; furthermore, when free from outside interference autonomous beings are capable of *acting* on those preferences to a considerable degree, thus living (to at least some extent) the life that they have chosen for themselves. (2019, 196)

Kagan holds that animals are significantly *less* autonomous than humans. One might think that this difference in degree of autonomy could be the basis for a hybrid view. To take such an approach, it might be argued that person status requires a degree of autonomy above a certain threshold, and that animals (usually or always) fall below that threshold while humans (usually or always) exceed it.

relevant non-moral facts, careful thought, and consideration of alternative views and arguments. Traditional surveys, including the sorts of surveys conducted by Caviola et al., are unable to distinguish between stable intuitions and superficial gut reactions that participants would retract after due reflection. For this reason, we think that alternative methodologies—such as a method known as deliberative polling—ought to be explored in this area.

Kagan's argument that animals are significantly less autonomous than humans relies on the claim that humans have long range plans, extending from "tomorrow, to next year, the next decade, and on and on to life's very end," and that humans have preferences concerning many different aspects of life "from the grand to the trivial"—for example, we have preferences about our careers and also have preferences about what to have for breakfast (2019, 197). By contrast, animals' preferences extend only to the near future and only concern a limited range of aspects of their lives. And this, Kagan thinks, means that animals are less autonomous than humans.

On Kagan's way of thinking, the magnitude of your autonomy increases (*ceteris paribus*) as your preferences extend further into the future and as the breadth and diversity of your preferences increase.¹³ But consider Monica, age 20, who correctly expects to live to age 80 and has a range of preferences concerning the remaining 60 years of her life (and has no preferences about what happens after she dies). As each year passes, Monica's anticipated years of future life diminish and the temporal distance over which her preferences extend correspondingly diminishes. Kagan's position implies that Monica becomes significantly less and less autonomous as she ages—even if she maintains full cognitive functioning until the end of her life. Similarly, consider Gabe, an aesthete with refined taste who has many detailed preferences, and Bob, a monk who wants only to fulfill a few religious duties each day and is indifferent regarding most other things. Kagan's position implies that Gabe is far more autonomous than Bob. These implications seem hard to defend, putting it mildly.¹⁴ So, we believe, Kagan has failed to persuasively argue that animals are less autonomous than humans.

These points show that there is some difficulty in producing an account of autonomy that renders animals significantly less autonomous than humans without

¹³ This is not to say that these are the *only* factors affecting magnitude of autonomy in Kagan's account. We're grateful to an anonymous reviewer for pushing us on this point.

¹⁴ Kagan responds to a problem that is related to, but less serious than, the problem we discuss here. Kagan allows that there is "normal variation" among humans and among animals. Among ordinary humans, for instance, he allows that some have a "somewhat better ability to plan" than others, that some have a "somewhat greater capacity for normative self-governance" than others, etc. (2019, 164). Given the prominent role that autonomy-related capacities play in Kagan's account of status-generating capacities (2019, 124–26), these differences mean that Kagan is committed to the result that, other things being equal, those ordinary humans have *slightly* higher moral status than others. This would seem to imply that, on Kagan's account, we should think of different ordinary people as having slightly different levels of moral status, which (to many) will seem objectionable. Kagan responds to this problem by making two points. First, he maintains that the differences between ordinary humans are typically so small as not to significantly affect how ordinary humans ought to treat one another in most real-world circumstances (2019, 166). Second, he defends a "practical realism" according to which "moral rules are to be evaluated with an eye toward our actual epistemic and motivational limitations" (2019, 292). Given this practical realism, he thinks, the best moral rules are likely to require us to treat all ordinary humans as if their moral status is exactly, not just approximately, equal. However, these points do not satisfactorily address our objection. As we have seen, Kagan's account supports the view that an ordinary human suffers a *dramatic* reduction in autonomy as they age and that ordinary humans can be *dramatically* more autonomous than others simply by having far more diverse and detailed preferences. Contrary to Kagan's response, then, his account does not support the view that differences in autonomy among ordinary humans are *typically* slight. And this in turn suggests that even given Kagan's practical realism, Kagan's account has difficulty avoiding the unsavory implication that we should treat some ordinary humans as if their status is dramatically lower than that of other ordinary humans.

carrying implausible implications (such as the implication that Monica becomes dramatically less autonomous as she ages or that Gabe the aesthete is dramatically more autonomous than Bob the monk). Kagan's account of autonomy fails to adequately resolve this problem. And we know of no alternative account that succeeds. Because of these difficulties, we want to explore what happens if it is assumed that animals and humans are equally autonomous. In the next section, we explore an autonomy-based rationale for the Hybrid Theory which is compatible with the claim that animals are as autonomous as humans.

3.2 The Value of Autonomy

Suppose for the sake of argument that animals and humans are equally autonomous. Then one might think that there is no way to argue that autonomy grounds the sort of difference between animals and humans that could support the Hybrid Theory. However, the defender of the Hybrid Theory still has at least two autonomy-based strategies up their sleeve, both based on the idea that animals' autonomy is not intrinsically morally valuable, whereas humans' autonomy is. If such an argument could work, then—as we will explain—it might support the Hybrid Theory (with the addition of a few further steps, which are themselves not uncontroversial). In the remainder of Sect. 3, we'll provide an outline of two such autonomy-based rationales for the Hybrid Theory.

Consider, first:

The First Capability Argument

(1a) Weak perspectivalism about goodness-for: X is intrinsically (i.e., non-instrumentally) good for an individual only if that individual can value X.

(2) Animals cannot value their own autonomy.

(3a) If animals' autonomy is not intrinsically good for them, then it is not intrinsically morally valuable.

Therefore, animals' autonomy is not intrinsically morally valuable.

We call (1a) 'weak perspectivalism about goodness-for' because it is weaker than an alternative premise, namely.

Strong perspectivalism about goodness-for: If an individual *does not* value X, then X is not intrinsically good for that individual.

Strong perspectivalism about goodness-for is implausible, as it implies that it would not be intrinsically good for a severely depressed person who does not value their own happiness to become happy. Weak perspectivalism avoids this problem: even a severely depressed person will typically be *able* to value their own happiness.

One objection to weak perspectivalism about goodness-for is suggested by Richard Kraut:

What is good for a plant is not a matter of what is good from the perspective of the plant, because it has no perspective, no outlook on the world, no goals, plans, or desires. That point by itself shows that the "for" of "good for S" is not perspectival when S is a plant. But it also shows that the "for" of "good for S" is never perspectival, for "good for S" does not have one meaning when S is a plant, and another when S is an animal or a human being. (2007, 94)

Arguably, plants cannot value their own survival or health (because plants do not have minds and valuing is arguably something only minds can do); yet, it might be maintained, survival and health are intrinsically good for plants. To respond, the weak perspectivalist about goodness-for has to argue either that: (i) plants are able to value survival and health despite being mindless (Rolston 1993); or (ii) either because of semantic ambiguity or context-sensitivity, the property attributed to plants using the phrase “good for” is different from the property attributed to animals and humans; or (iii) nothing is intrinsically good for plants (Feinberg 1974). Weak perspectivalists about goodness-for have some work to do.

An alternative that sidesteps this problem for the First Capability Argument would be to reject weak perspectivalism about goodness-for while maintaining that something’s being intrinsically good for an individual doesn’t, by itself, entail that that thing is intrinsically morally valuable; some additional requirement must be met (Streiffer 2003, 80–87; Belshaw 2016). One possible requirement would be weak perspectivalism about intrinsic moral value: something’s being intrinsically good for an individual makes that thing intrinsically morally valuable only if that individual can value that thing. On this view, things can be intrinsically good for plants, but because plants cannot value those things, those things are not thereby intrinsically morally valuable. This suggests the following:

The Second Capability Argument

(1b) Weak perspectivalism about intrinsic moral value: X is intrinsically morally valuable because X is intrinsically good for an individual only if that individual can value X.

(2) Animals cannot value their own autonomy.

(3b) If animals’ autonomy is not intrinsically morally valuable because it is intrinsically good for them, then it is not intrinsically morally valuable.

Therefore, animals’ autonomy is not intrinsically morally valuable.

Even if weak perspectivalism about goodness-for or about intrinsic moral value could be successfully defended, the remaining premises of the Capability Arguments also need defense.

To argue for (2) one might argue that (i) one cannot value something unless one can conceive of it, and (ii) one cannot conceive of autonomy unless one possesses certain linguistic abilities that animals lack.

To make such an argument, one might note, first, the ways in which humans seem to value their autonomy: they work to cultivate their autonomy (e.g., by trying to improve their own self-control, or by training their children to have self-control), and they object to paternalism that limits their autonomy even when they know the limitation is in their own overall best interest. Valuing autonomy in these ways seems to require the ability to *conceive* of autonomy. In contrast, animals do not seem to value autonomy. This, it might be argued, is because they are not even capable of conceiving of autonomy (perhaps because autonomy is imperceptible and thus requires certain linguistic abilities for its conceptualization which animals lack).

To argue for either (3a) or (3b), one would need to argue against a range of different ways that animals’ autonomy might acquire intrinsic moral value on grounds other than its being intrinsically good for the animals. For example, on some views, what is intrinsically good for something beautiful is intrinsically morally valuable

even when it is not intrinsically good for anyone (such that, e.g., you have moral reasons to refrain from destroying something beautiful even if that beauty will never be enjoyed by anyone). Animals' autonomy might conceivably have moral value in a similar way; those who want to develop either of the Capability Arguments will need to argue against such views.

Defending all three of the premises of either of the Capability Arguments will clearly be a major task. We've offered some brief suggestions about how one might try to go about such a task. Suppose that it can be established, via one of the Capability Arguments, that animals' autonomy is not intrinsically morally valuable. What then?

In the next stage of reasoning for the Hybrid Theory, it might be argued that humans *are* capable of valuing their own autonomy (for they have the conceptual-linguistic abilities needed to do so), and that this means that humans' autonomy *is* intrinsically morally valuable. On such a view, it would not (or should not) be claimed that animals' lives utterly lack moral value. Animals value companionship, food, play, and so on, and it is highly plausible that these aspects of animals' lives have intrinsic moral value. And when these things feature in humans' lives, they are no less intrinsically morally valuable than when they feature in animals' lives. Thus, the picture that emerges is one in which humans' and animals' lives are intrinsically morally valuable in many of the same ways; but humans' lives have at least one additional element of moral value, namely autonomy.

3.3 Value Conservatism and the Hybrid Theory

Next, it can be argued that some types of moral value are *morally fungible*, and others are not. A type of moral value is morally fungible when it is always permissible (*ceteris paribus*) to optimifically destroy instances of that type of value. Morally fungible value is the sort of value that never *stands in the way* of optimific action.

Here it is useful to consider a view proposed by G. A. Cohen:

I claim that we devalue the valuable things we have if we keep them only so long as nothing even slightly more valuable comes along. Valuable things command a certain loyalty. If an existing thing has intrinsic value, then we have reason to regret its destruction *as such*, a reason that we would not have if we cared only about the value that the thing carries or instantiates. My thesis is that it is rational and right to have such a bias in favor of existing value, that, for example, if you happily replace a fine statue by a merely somewhat better one, the production of which requires destruction of the original statue, then you mistreat the now destroyed work as (so to speak) having had the merely instrumental value of being a vessel of aesthetic value. (2011, 210)

Contra Cohen, destroying a valuable object to produce an object of greater value does not necessarily involve treating the first object as having *merely* instrumental value.¹⁵

¹⁵ You might simply have thought the original statue was valuable intrinsically as a beautiful work of art and yet *less* valuable than the better statue that could be made from it. See (Chappell 2015) for useful

But Cohen's position can be separated from his reasoning for it and his position is worth taking seriously. Cohen's position, call it *value conservatism*, is that intrinsically valuable objects call for preservation and protection such that we mistreat them if we destroy them merely to pursue greater value.

Cohen's example involves a statue's *aesthetic* value. Problems emerge for Cohen's position when it is construed as a *moral* thesis. Even if the statue has a high degree of intrinsic value, many will say, it is still *morally* acceptable, provided all else is equal, to destroy the statue to obtain an object of even greater value.

Yet Cohen's position as a moral thesis when applied to *human beings* will be far less controversial. Many people think that, even if the value inherent in beautiful statues is fungible, the particular kind of value inherent in human beings is not—and therefore it is at least sometimes wrong to kill one human being to save many others even if doing so is optimific. This suggests a *limited* value conservatism, according to which some sorts of intrinsic moral value are fungible while others are not.

Given such a view, the defender of the Hybrid Theory could argue that autonomy's moral value is a form of non-fungible value, whereas other sorts of moral value—in particular, the sorts of moral values that figure in animals' lives, such as the moral value of companionship, enjoyment of food, play, and so on—are fungible. If this could be established, then it would go some distance toward defending the Hybrid Theory.

When an individual is killed, they are destroyed—but their *properties* are also, in an important sense, destroyed as well. When an autonomous being is destroyed, their autonomy is destroyed; when a happy being is destroyed, their happiness is destroyed; and so on. Now, we are supposing that pigs and humans are equally autonomous. But if the defender of a hybrid view can show (via one of the Capability Arguments) that pigs' autonomy is not intrinsically morally valuable whereas humans' autonomy is intrinsically morally valuable, and if it can also be shown that the sort of moral value that attaches to humans' autonomy is non-fungible, then it could be argued that killing a human being—thus destroying their autonomy—destroys something of non-fungible moral value, whereas killing a pig only destroys what is of fungible moral value. And this, in turn, could explain why in a version of Outbreak where one pig must be fatally infected with the virus to save five pigs—call this Pig Outbreak—it is permissible to kill the one to save the five, whereas in a version where one human must be fatally infected to save five humans—call this Human Outbreak—it is not permissible to kill the one to save the five.

Those who want to make this type of argument for the Hybrid Theory will need to argue that when autonomy is intrinsically morally valuable, its moral value is non-fungible whereas other sorts of moral values—specifically, the sorts of moral values that figure in the lives of animals deemed persons—are fungible. That's a difficult challenge: friendship, for example, is something the moral value of which might be believed to be non-fungible; and it seems that pigs and other animals can have friends. If that difficult challenge can't be met, then the present strategy for defending the Hybrid Theory can't succeed. This type of argument for the Hybrid Theory also requires that one of the Capability Arguments be successful. So, an autonomy-based

related discussion.

argument for the Hybrid Theory of the sort we've sketched here will require additional work; but we hope that the outline we've given is sufficient to show that this line of argument merits consideration and further investigation.

4 Commitment Utilitarianism and the Hybrid Theory

As we observed above, defenders of the Hybrid Theory need to identify a difference between *dersons* and *uersons* and explain how this difference can explain why the beings they've classified as *dersons* (humans, according to typical hybrid views) are in fact *dersons* and why the beings classified as *uersons* (animals, again according to typical hybrid views) are in fact *uersons*. In this section, we investigate whether differences regarding the capacity for consent can make the needed difference. We'll outline a position that we'll call commitment utilitarianism (4.1), explain how commitment utilitarianism might be used to defend the Hybrid Theory (4.2), and discuss objections (4.3).

4.1 Commitment Utilitarianism

A cake needs to be divided between Mother Teresa and George Constanza. Suppose the optimific action is to give precisely two-thirds of the cake to Teresa and one-third to George (because Teresa likes cake more than George, but Teresa's utility function exhibits diminishing marginal returns to cake). Teresa knows this, but because Teresa is so saintly, she wants all of the cake to go to George; and George, who is far more self-interested than Teresa, is fully happy with that arrangement as well. Although it's not implausible to think that Teresa is *entitled* to two-thirds of the cake it nevertheless seems mistaken to think that Teresa must *receive and accept* that amount of cake. In fact, if Teresa wants all of the cake to go to George, then it seems permissible and perhaps even obligatory to give the cake to George.

Importantly, these intuitive judgments seem plausible even from—perhaps especially from—a utilitarian perspective. After all, a core idea of utilitarianism is that everyone is entitled to equal consideration of interests. If we accept this idea, then it is plausible to think that Teresa is entitled to two-thirds of the cake, since that is the verdict that equal consideration of interests supports. We might say that two-thirds of the cake is Teresa's *utilitarian due*. But all of this is consistent with and perhaps supportive of the thought that Teresa can do what she wishes with her utilitarian due, including relinquish it and give it to George.

These considerations suggest the following idea. According to *consent utilitarianism*, everyone is entitled to be treated at least as well as they would be treated were their interests to be counted in the normal utilitarian style; but it can be permissible to treat someone worse than that if they give their consent.¹⁶

¹⁶ A utilitarian might try to handle the Teresa and George case by understanding individuals' interests in terms of their preferences. But this solution will be costly for those who believe—as many utilitarians do believe—that our interests *cannot* be spelled out in terms of our preferences. Further, preference utilitarianism might not do full justice to our intuitions about the Teresa and George case. After all, preference utilitarianism takes account of the *full range* of an individual's preferences. So we can imagine a case

In an unpublished manuscript, Christopher Meacham (n.d.) has (independently) hit on the idea of consent utilitarianism. Meacham observes that many philosophers in the utilitarian tradition are attracted to a certain self-other asymmetry. On the one hand, it seems permissible for one to make large sacrifices to provide relatively small benefits to others: for instance, a parent can voluntarily work two grueling weeks of overtime to provide a birthday present for her child, even if the benefit to the child will be small in comparison to the pain of the overtime. On the other hand, it seems impermissible for one to force others to make large sacrifices to provide relatively small benefits for oneself.

Versions of utilitarianism have been devised to accommodate this asymmetry. However, Meacham observes that our intuitive judgments in this neighborhood don't always track the self-other divide. For example, it seems permissible for you not to talk your spouse out of working two weeks of overtime to provide a birthday present for your and your spouse's child, even if the costs to your spouse outweigh the benefits to your child. So, Meacham proposes that the relevant asymmetry has to do with *consent* rather than the self-other divide, and this leads him to propose a view quite like consent utilitarianism (n.d.).

However, consent utilitarianism runs into difficulties. Consider cases of extreme self-sacrifice. Suppose you volunteer to give both of your legs (a major loss to you) to save George's middle toe (a relatively minor loss to him). It seems that the mere fact that you have consented to engage in this sort of *lopsided altruism*, as we might call it, is not sufficient to show that it is permissible for you to do so (if we think that there are duties to the self) or that it is permissible for others to help you.

However, when we think about cases where an individual is deeply committed to being lopsidedly altruistic, our intuitions change. For example, in light of Mother Teresa's deep commitment to extreme forms of altruistic self-sacrifice, it seems permissible to let her make significant sacrifices for comparatively small benefits to others, and permissible to help her do so.

The intuitions we've just reported suggest that consent utilitarianism should be amended: (i) by default, everyone is entitled to be treated as well as they would be treated if we acted optimifically; and (ii) we can permissibly treat people worse than that if and only if they have a suitably deep commitment that prescribes that they be so treated. Call this view *commitment utilitarianism*. Commitment utilitarianism avoids the counterintuitive implication of consent utilitarianism that an ordinary person's consent licenses you to take that person's legs to save George's middle toe. Rather, you can take someone's legs to save George's middle toe only if they have a Mother Teresa-level of commitment to such extreme altruism. Commitment utilitarianism's verdict here seems more plausible.¹⁷

where Teresa says that the cake should go to George even though the full range of preferences would be more completely satisfied by giving the cake to Teresa instead. In such a case Teresa's willingness to give the cake to George still seems to justify giving the cake to George.

¹⁷ Analogies with other, non-altruism cases also support the claim that commitment utilitarianism's verdict is more plausible. Consider gender affirmation surgery, or any other major plastic surgery that is largely irreversible. Many believe that mere consent is not enough to justify a doctor in proceeding with such surgery; rather, doctors should look for evidence that patients have a deep commitment to having such surgery before proceeding.

4.2 From Commitment Utilitarianism to the Hybrid Theory

To see the move from commitment utilitarianism to the Hybrid Theory, let's return to Pig Outbreak and Human Outbreak. For the commitment utilitarian, each of the five pigs in Pig Outbreak is entitled to have their life saved, because their life will be saved if the agent does what's optimific. Being saved is the five pigs' utilitarian due. Therefore, it is impermissible for you to allow them to die unless they have some sort of deep commitment that prescribes that they be so treated. It's fair to say that pigs lack any such deep commitment. So, according to commitment utilitarianism, it is not acceptable to allow the five pigs to die and the agent should therefore kill the one.

What about Human Outbreak? It may initially seem implausible that anyone apart from extreme altruists has a deep commitment that prescribes that they be allowed to die in this sort of case. But such a commitment may be fairly widespread even though extreme altruism is not at all widespread.

Let's first discuss the notion of a *deep* commitment as it figures in commitment utilitarianism. We should say a commitment is deep when it stands in relationships of mutual support with the subject's other beliefs and commitments.¹⁸ This is because it seems plausible to think that commitments that are not deep in this sense cannot justify sub-optimific treatment, whereas commitments that are deep in this sense plausibly can justify such treatment. To support this, consider the Mother Teresa case again: voluntarily being treated sub-optimifically for others' benefit is tightly integrated into Mother Teresa's worldview, a worldview that takes the Christian ideal of self-sacrifice unusually seriously. And it is because of this, one might propose, that it seems acceptable and perhaps even required to allow her to submit to such treatment, even where such treatment is highly costly to her.

As we've already mentioned, almost no one has a Mother Teresa-style commitment to extreme altruism. But most of us seem to have a series of deontological commitments. This is illustrated by the fact that when presented with generic versions of Human Outbreak, we tend to say that it is wrong to kill the one. And it seems plausible that these judgments reflect deontological commitments that (i) are highly coherent with our other beliefs and commitments and thus count as deep in the requisite way and (ii) prescribe that the five should not be saved in such cases and so, *a fortiori*, prescribe that the five should not be saved in such cases even when we happen to be among the five.

But not *everyone* explicitly endorses deontological judgments about these sorts of cases. There are some utilitarians, after all, who will disavow any deep commitment to deontological judgments. If such utilitarians truly do lack deep deontological commitments, then we can't appeal to their deep commitments to justify allowing them to die in a case like Human Outbreak. This conclusion won't satisfy a typical defender of the Hybrid Theory, who will want to say that even utilitarians count as persons.

¹⁸ Thus understood, the depth of a commitment comes in degrees. This opens up the possibility of a more complicated version of commitment utilitarianism: If an action X treats an individual Y in a detrimental way and X is sub-optimific, then the more harmful X is to Y, the deeper Y's commitment to being so treated needs to be for X to be justified.

What is needed for the present purpose, then, is an argument that all or nearly all human beings are deeply committed to deontological judgments. To make such an argument, one could argue that everyone who views themselves as ends in themselves is deeply committed to such deontological judgments. If this view is correct, then it would arguably apply to a very large number of humans, since it is arguable that a great many humans see themselves as ends in themselves. It may be that viewing oneself in such a way is implicit in viewing oneself as a rational agent, as Kant seems to have thought (Kant 1785, 37 [4:429]). Christine Korsgaard claims that every human being views themselves as ends in themselves:

After all, every human being pursues the things that are important to himself and to those whom he loves as if they were important *absolutely*.... And just by doing that, we claim our own standing as ends in ourselves. For when we claim that the things that are important to us should be treated as important absolutely, just because they are important to us, we also claim that we are important ourselves. (2012; See also 2018, 92–95)

If Korsgaard's claims are even approximately correct, and if those claims are conjoined with the Kantian view that everyone who views themselves as ends in themselves is deeply committed to deontological judgments, then we would have an argument for the view that nearly *all* humans are deeply committed to deontological judgments.

Here's a summary of such a Kantian argument. If you see yourself as an end in yourself, then you are committed to being opposed to being treated as a mere means. And when the one is killed in Human Outbreak, the one is treated as a mere means to another's end. An intermediate conclusion then follows: If you see yourself as an end in yourself, then you are committed to being opposed to being killed to save the five in Human Outbreak.

The argument continues: If, as someone who views themselves as an end in themselves, you are committed to being opposed to being killed in order to save the five in Human Outbreak, then you must on pain of inconsistency oppose the killing of *others* who view themselves in the same way if they are the one in Human Outbreak. It then follows that, if you regard yourself as an end in yourself, you are committed to being opposed to killing *anyone* who views themselves as ends in themselves—even if you are among the five in Human Outbreak.

And such a commitment is arguably deep, because our view of ourselves as ends in ourselves is arguably central to our worldview. Thus, we have the makings of an argument that most or all ordinary people are deeply committed to being allowed to die if they are among the five in Human Outbreak. And this means, according to commitment utilitarianism, that we can justify allowing the five to die in Human Outbreak, even though this is sub-optimal, provided the five are ordinary people.

Now, the typical defender of the Hybrid Theory is going to want to say a bit more than that. They will want to say that allowing the five to die in Human Outbreak is *obligatory*, not just that it can be justified. They will want to say that there is a deontological constraint against killing one human being to save five in such cases. To argue for that deontological constraint, the defender of the Hybrid Theory can say that if

one has a deep commitment that prescribes that they be treated in a certain way, this provides a reason of some strength in favor of treating them in that way. If this type of reason is strong enough to outweigh other countervailing reasons that may be present in cases like Human Outbreak, then one can argue that the agent in cases like Human Outbreak has most reason to allow the five to die. And it is plausible that we at least typically are obligated to do what we have most reason to do. Thus, one may argue that commitment utilitarianism implies that it is not only permissible but obligatory to allow the five to die in cases like Human Outbreak.

No similar argument exists for allowing the five pigs to die in cases like Pig Outbreak because pigs, unlike humans, do not see themselves as ends in themselves.¹⁹ Thus, an asymmetry emerges.

This asymmetry won't perfectly trace the human/animal divide. Some humans—e.g., severely mentally disabled people or people who have highly unusual ideas about themselves—are arguably not deeply committed to viewing themselves as ends in themselves, so we might end up with the result that it is permissible or required to kill them whenever doing so is optimistic. Similarly, perhaps some animals are deeply committed to viewing themselves as ends in themselves, and thus might end up being protected by deontological constraints.

But this asymmetry *could* be used to draw the person/person divide. For according to the hybrid view that emerges from this line of reasoning, to be a person *just is* to be the sort of individual who sees themselves as an end in themselves. And so we have arrived by a new path at a position that supports the kind of person/person asymmetry endorsed by the Hybrid Theory: an asymmetry in which those who qualify as persons (namely, humans and animals (if any) who see themselves as ends in themselves) are protected by deontological constraints whereas those who qualify as persons (namely, animals and others who do not see themselves as ends in themselves) aren't.

4.3 A Few Objections

The view emerging from these considerations is not without serious difficulties. Here we'll mention three.

First, according to commitment utilitarianism, deep commitment can justify sub-optimistic treatment. But this sort of deep commitment can exist even in the absence of actual consent. Take Human Outbreak. The argument we are developing here is that each of the five are deeply committed to a deontological judgment according to which it is wrong to kill the one to save the five, and therefore they are deeply committed to the judgment that they must be allowed to die.

But suppose the five are all begging to be saved. And suppose the agent calls back to them, "Sorry, I can't help you! Since you regard yourselves and the one as well as ends in yourselves, you are deeply committed to being allowed to die in these circumstances. And this deep commitment gives me a reason that is sufficiently strong that

¹⁹ We think this is a plausible assumption, because we think that seeing oneself as an end in itself requires a kind of abstract thought of which pigs seem incapable. It appears Korsgaard, in some of her publications, disagrees (2012, 368–9), but we do not have the space to argue for this assumption here. We thank a reviewer for pushing us to flag this.

I am in fact *obligated* to allow you to die, no matter how loudly you protest.” Such a response may seem a bit weak. Why should respecting their deep commitments take priority over what they say they want, especially in circumstances where honoring their deep commitments results in the death of the very people whose deep commitments we are (allegedly) honoring?

One response would be to point out that commitment utilitarianism has partners in crime. Consider the problem of justifying punishment of criminals. Many philosophers have argued that punishment can be justified by a commitment grounded in hypothetical consent: in ideal circumstances, the state of nature, or some such hypothetical state, the criminal *would* consent to be bound by laws and therefore to be subject to punishment if found guilty of violating those laws, and this hypothetical consent justifies punishment of the criminal even in cases where actual consent to punishment is vigorously refused. This idea resembles the commitment utilitarian’s idea that deep commitments can justify treating people even in ways that they explicitly resist in the moment. So, perhaps the commitment utilitarian can argue by analogy with already-developed maneuvers that appeal to hypothetical consent.

But attempts at justifying coercion and other forms of non-consensual interactions in terms of hypothetical consent are themselves controversial. Ronald Dworkin, for example, notes that it doesn’t follow from the fact that I would have agreed to play a game by different rules that it is fair to enforce those rules against me even if I have not agreed (1973, 501). A similar objection to commitment utilitarianism might be available as well.

A second obvious problem is that the argument developed here relies on a Kantian chain of reasoning that one might be reasonably skeptical about, especially the claim that killing the one requires treating the one merely as a means. Why not say, as a utilitarian would, that it merely requires treating the one as an end whose importance is outweighed by the importance of the five? A defender of the Hybrid Theory who wants to use the strategy outlined here needs a response to this point.

A third problem is that the argument produces implausible results in some cases. Consider the following case:

Mixed Push: You can push a human being into the way of a trolley to save N pigs, where N is a number of pigs such that pushing the human into the way of the trolley is optimistic.

The pigs are not deeply committed to relinquishing their utilitarian due; therefore, according to the argument developed here, the human has to be pushed into the way of the trolley. This is counterintuitive. And most of those who like a hybrid view will not want to sign on for this idea. Thus, it appears that the argument for the Hybrid Theory that we’ve been exploring cannot get us all the way to a fully intuitively acceptable version.

This last issue is a symptom of a more general feature (or bug) of the rationale we are discussing here. Specifically, the rationale relies principally on *commitments* of those who stand to benefit from *violation* of deontological constraints; it does not principally rely on morally important characteristics (e.g., non-fungible moral value) of those who stand to benefit from *respect* for deontological constraints.

5 Conclusion: Can Relationships Provide a Basis for the Hybrid Theory?

We have so far considered three arguments for the Hybrid Theory: one case-based, one focused on the non-fungible moral value of autonomy, and one focused on the capacity for deep commitments. Although case-based arguments for the Hybrid Theory have appeared elsewhere in the animal ethics literature, the two rationales considered in Sect. 3 and Sect. 4 are new (to our knowledge). We think all three of the lines of reasoning we've discussed deserve further consideration, but all face serious challenges and it is quite possible that none of them will ultimately be successful. In that case, those who are intuitively drawn to the Hybrid Theory will have to develop alternatives or give up their view.

A further line of reasoning that leads in the direction of the Hybrid Theory proceeds from observations about the relationships that human beings have with one another and the differences between those relationships and the relationships which animals are capable of entering into. In this concluding section, we'll briefly discuss the prospects for this sort of strategy.

According to a view we will call *relationalism*, relationships have moral significance: relationships can generate special obligations. According to a standard version of relationalism, we have special obligations to our family members, professional colleagues, and fellow citizens, among others (Ross 1930, esp. 19). Different versions of relationalism will spell out different ways of thinking about these special obligations. A typical relationalist view is that the interests of your family members, colleagues, and fellow citizens have extra weight for you and you are obligated to act accordingly. Thus, for example, if you have to choose whether to save your spouse or a stranger from drowning, you are obligated to save your spouse, even if the stranger has a slightly greater interest in being saved.

Relationalism offers some resources to deontologists: deontologists believe we are sometimes obligated to act sub-optimifically, and relationalism entails precisely that. A further question, however, is whether a complete deontological theory can be fully fleshed out in relationalist terms. The project of constructing a fully relationalist deontology may seem to be a non-starter because there may seem to be some deontological verdicts that cannot be supported on relationalist grounds.

To see this, consider Human Outbreak. Even though it is optimific to kill the one to save the five, deontologists typically say that you must not do so. But how can this verdict be explained in relationalist terms? In typical versions of the case, all of these individuals are strangers to you, and so you have no relationships with any of them.

But relationalists have a way to explain how it can be wrong to kill the one. Relationalists can argue that the very act of killing creates a relationship. The one is, in the moment of killing, no longer a stranger; rather, the one becomes a *victim* of violence committed by the agent. This relationship, it can be argued, has moral significance, no less (and perhaps in some cases, more) than the relationships between family members, colleagues, and fellow citizens. The relationalist can maintain that it is in virtue of this fact that killing the one is morally wrong. Further, the relationalist can argue that, when the agent allows the five to die, no comparably significant relationship with the five is formed. Thus, the relationalist can argue that there is a moral

asymmetry between letting die, on the one hand, and killing, on the other, and can use this asymmetry to explain why killing the one to save the five is morally worse than letting the five die.

These points show that relationalists might have the resources to explain not only why we have special obligations to family members, colleagues, citizens, and so on, but also to explain why there are moral constraints against optimific harm, even where those moral constraints apply to individuals who are (prior to the action in question) total strangers. The project of spelling out a fully satisfactory relationalist deontology may be more promising than it at first appears.

A relationalist account might look like a hybrid view if the relationalist argues that animals (or most animals) lack the ability to enter into relationships with others, including humans, that are directly morally significant. To make such an argument, it would be necessary to identify a feature of animals that prevents them from entering into such relationships. If such a feature could be identified, then the relationalist could argue that we have no special obligations to animals (because we have no relationships with them that could generate such obligations) and that we have no moral constraints against optimific harm to animals, or, at least, no such moral constraints arising from our duties to the animals themselves. Nevertheless, it might still be maintained that animals' interests provide morally significant reasons for action and must be taken into account and given appropriate weight.

Of course, this sort of view requires that animals lack the ability to enter into morally significant relationships with others. How could one defend such a claim? Here is a possible strategy:

The Third Capability Argument

(1) Weak perspectivalism about relationships: An individual can be in a morally significant relationship only if they are capable of understanding that they are in a morally significant relationship.

(2) Animals are not capable of understanding that they are in morally significant relationships.

Therefore, animals cannot be in morally significant relationships.

Those who want to make this argument have some work to do. For one thing, weak perspectivalism about relationships has several unsavory implications: for example, it implies that any human being who cannot understand that they are in morally significant relationships cannot be in such relationships. This would seem to imply, for example, that there is no morally significant relationship between a newborn baby and its parents. If one tries to avoid this conclusion by arguing that the cognitive capacities required to recognize that one is in a morally significant relationship are so meager that even newborns have that capacity, then it will, of course, be difficult to argue that non-human animals do not have the same ability, thus making the second premise of the argument difficult to defend. Another problem is that many people view themselves as being in morally significant relationships with their pets and other animals under their care which give rise to special obligations (Burgess-Jackson 1998). This belief is flatly inconsistent with the very general conclusion of the Third Capability Argument.

If the Third Capability Argument is difficult to defend, and if no other way of arguing that animals cannot be in morally significant relationships surfaces, then one may

need to concede that animals can, at least in principle, be in morally significant relationships with humans. But it could still be maintained that human-animal relationships typically differ greatly from human-human relationships. It could be argued, for example, that the moral significance of a spousal relationship between two human beings differs dramatically from the relationship that any human being has with any animal. And this, in turn, could support a variety of different views about the differences between our typical moral obligations to animals and our typical moral obligations to one another.

Would the resulting relationalist view count as a true hybrid view? According to the relationalist view, what explains the difference between our obligations to persons and our obligations to animals is that we can enter into special relationships with the latter but not with the former. The core idea of the Hybrid Theory is that persons and animals differ in terms of their moral status. On one way of thinking about moral status, obligations that result from special relationships are not constitutive of or even relevant to moral status. Such a view would readily explain the plausible thought that the fact that you are obligated to save your spouse rather than a stranger does not imply that your spouse has a higher moral status than the stranger. On this sort of view, the relationalist is *not* saying that two classes of individuals differ in terms of their moral status, and thus does not count as a hybrid view, strictly speaking. However, if it could be argued that facts about moral status could be reduced to facts about special relationships, then the relationalist view could count as a hybrid view. We leave this possibility open for further exploration in the future.

References

- Belshaw, C. (2016). "Death, Pain, and Animal Life." In *The Ethics of Killing Animals*, edited by Tatjana Višak and Robert Garner, 32–50. Oxford: Oxford University Press
- Burgess-Jackson, K. (1998). Doing Right by Our Animal Companions. *The Journal of Ethics*, 2(2), 159–185
- Caviola, L., Kahane, G., Everett, J. A. C., Teperman, E., Savulescu, J., & Faber, N. S. (2021). Utilitarianism for Animals, Kantianism for People? Harming Animals and Humans for the Greater Good. *Journal of Experimental Psychology: General*, 150(5), 1008–1039
- Chappell, R. Y. (2015). "Value Receptacles." *Noûs* 49 (2):322–32
- Cohen, C. (1997). "Do Animals Have Rights?" *Ethics & Behavior*, 7(2), 91–102
- Cohen, C. (1986). The Case for the Use of Animals in Biomedical Research. *The New England Journal of Medicine*, 315(14), 865–870
- Cohen, C., & Regan, T. (2001). *The Animal Rights Debate*. Lanham, MD: Rowman & Littlefield
- Cohen, G. A. (2011). "Rescuing Conservatism: A Defense of Existing Value." In *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*, edited by R. Jay Wallace, Rahul Kumar, and Samuel Freeman, 203–30. Oxford: Oxford University Press
- Dworkin, R. (1973). The Original Position. *The University of Chicago Law Review*, 40(3), 500–533
- Feinberg, J. (1974). "The Rights of Animals and Unborn Generations." In Athens, GA, William Balckstone (Eds.) *Philosophy and Environmental Crisis* (pp. 43–68). University of Georgia Press
- Kagan, S. (2019). *How to Count Animals, More or Less*. Oxford: Oxford University Press
- Kahane, G., & Caviola, L. n.d. "Are the Folk Utilitarians about Animals?" *Philosophical Studies*
- Kant, I. (1785). *Groundwork of the Metaphysics of Morals*. Edited and translated by Mary Gregor. Cambridge: Cambridge University Press
- Killoren, D., and Robert Streiffer (2020). Utilitarianism about Animals and the Moral Significance of Use. *Philosophical Studies*, 177(4), 1043–1063

- Korsgaard, C. (2012). "Getting Animals in View." *The Point Magazine*. December 23, 2012. <https://the-pointmag.com/examined-life/getting-animals-view/>
- . *Fellow Creatures: Kantian Ethics and Our Duties to Animals*. Oxford: Oxford University Press
- Kraut, R. (2007). *What Is Good and Why: The Ethics of Well-Being*. Cambridge, MA: Harvard University Press
- Lomasky, L. (2013). "Is It Wrong to Eat Animals?" *Social Philosophy & Policy*, 30(1–2), 177
- Meacham, Christopher. n.d. "Utilitarianism, Altruism, and Consent." Unpublished Manuscript
- Mill, J. S. (1871). *On Liberty, Utilitarianism, and Other Essays*. Edited by Mark Philp and Frederick Rosen. Oxford: Oxford University Press
- Nozick, R. (1974). *Anarchy, State, and Utopia*. Oxford: Basic Books
- Regan, T. (1983). *The Case for Animal Rights*. Berkeley: University of California Press
- Ringach, D. L. (2011). The Use of Nonhuman Animals in Biomedical Research. *The American Journal of the Medical Sciences*, 342(4), 305–313
- Rolston, Holmes, I. I. I. (1993). "Value in Nature and the Nature of Value." In Robin Attfield and Andrew Belsey (Eds.) *Philosophy and the Natural Environment (Royal Institute of Philosophy Supplement: 36)* (pp. 13–30). Cambridge: Cambridge University Press
- Ross, D. (1930). *The Right and the Good*. Edited by Philip Stratton-Lake. Oxford: Clarendon Press
- Sidgwick, H. (1907). *The Methods of Ethics*. London: Palgrave Macmillan: Seventh
- Streiffer, R. (2003). *Moral Relativism and Reasons for Action*. New York, NY: Routledge: Studies in Ethics
- Thomson, J. J. (1990). *The Realm of Rights*. Cambridge: Harvard University Press
- Varner, G. E. (2012). *Personhood, Ethics, and Animal Cognition: Situating Animals in Hare's Two-Level Utilitarianism*. Oxford University Press

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.