Abstract

I ask how we should treat individuals in cases of uncertainty about whether or not they are sentient (on the assumption that sentience is necessary and sufficient for moral status). I present and evaluate three options. First, we can follow an incautionary principle that permits us to treat individuals as non-sentient in cases of uncertainty. Second, we can follow a precautionary principle that requires us to treat individuals as sentient in cases of uncertainty. Third, we can follow an expected value principle that requires us to multiply our credence that individuals are sentient by the amount of moral value that they would have if they were. I then draw two conclusions. First, we should reject the incautionary principle. Second, if we accept either the precautionary principle or the expected value principle, then it will follow that we morally ought to treat many individuals, such as invertebrates, as having at least partial moral status.
1. Introduction

In 2003 David Foster Wallace took a trip to Maine to write an article for *Gourmet Magazine* about what it was like to attend the Maine Lobster Festival. But the article that he ended up publishing, titled “Consider the Lobster,” is not so much a travelogue as a deeply personal examination of whether or not we have a moral duty not to kill lobsters for food.¹ About halfway through this article, Wallace puts his finger on a problem that, I think, deserves much more philosophical attention than it currently receives. He writes:

[T]he questions of whether and how different kinds of animals feel pain ... turn out to be extremely complex and difficult. And comparative neuroanatomy is only part of the problem. … [The] principles by which we can infer that others experience pain ... involve hard-core philosophy – metaphysics, epistemology, value theory. … And everything gets progressively more abstract and convoluted as we move farther and farther out from the higher-type mammals into cattle and swine and dogs and cats and rodents, and then birds and fish, and finally invertebrates like lobsters.²

The problem that Wallace is describing here, which I will call the *sentience problem*, is this: Suppose that (a) an individual has moral status if and only if they are sentient and (b) we are not always certain which individuals are sentient. How should we treat individuals in cases of uncertainty about whether or not they are sentient?

Wallace does not try to answer this question in his essay. Instead, he argues that we have at least some reason to think that lobsters do, in fact, consciously experience pleasure and pain,

---
and he concludes by expressing confusion about what to do in light of this situation. The philosophical literature is similar, though somewhat less candid. Many philosophers are aware of the sentience problem, but few discuss it in any detail, let alone try to solve it. Instead, most philosophers who regard sentience as morally relevant try to get around this problem in one of two ways. First, they try to argue that particular animals are, or are not, sentient. Second, insofar as they still feel uncertain about particular animals, they simply stipulate a solution to the sentience problem without saying much about why we should accept it.

My aim in this paper is to present and evaluate three possible solutions to the sentience problem. First, we can follow an incautionary principle that permits us to treat individuals as non-sentient in cases of uncertainty. Second, we can follow a precautionary principle that requires us to treat individuals as sentient in cases of uncertainty. Finally, we can follow an expected value principle that requires us to multiply our credence that individuals are sentient by the amount of moral value that they would have if they were. I will not try to say which one of these principles is correct in this paper. Instead, I will argue for two preliminary conclusions. First, we should reject the incautionary principle. Second, if we accept either the precautionary principle, the expected value principle, or both, then we morally ought to treat many individuals, such as invertebrates, as having at least partial moral status. And if this conclusion holds, then morality involves much more cluelessness and demandingness than we might have expected or hoped.

---

3 For a notable exception, see Alex Guerrero, “Don’t Know, Don’t Kill: Moral Ignorance, Culpability, and Caution,” Philosophical Studies 136:59–97, 2007. Guerrero argues for a principle called “Don’t Know Don’t Kill,” according to which we should not kill individuals unnecessarily in cases of uncertainty about whether or not they have moral status. My evaluation of the precautionary principle will double as an evaluation of DKDK.

4 See, for example, David DeGrazia, Taking Animals Seriously: Mental Life and Moral Status (Cambridge: Cambridge University Press, 1996).

I will proceed as follows. In section 2, I will present the background assumptions in moral philosophy and philosophy of mind that motivate the sentience problem. In sections 3-5, I will present and evaluate the incautionary principle, precautionary principle, and expected value principle, respectively. Finally, in section 6, I will draw my preliminary conclusions.

2. Assumptions and Clarifications

I begin by presenting the assumptions in moral philosophy and philosophy of mind that motivate the sentience problem, along with a few clarifications and qualifications about my discussion in this paper.

The first assumption that we need to make in order to motivate the sentience problem is **sentientism about moral status**. On this assumption, an individual has moral status, i.e. matters morally for their own sake, if and only if they are sentient, i.e. capable of consciously experiencing pleasure or pain. Note that the kind of consciousness that matters for our purposes here is *phenomenal* consciousness. That is, on the theory of moral status that we will be considering, your experiences count as conscious in the morally relevant sense if and only if there is “something that it is like” for you to be having them.\(^6\) As we will see, this requirement that you have a private, subjective, qualitative feeling corresponding to your pleasure and pain experience is part of what makes the sentience problem, as defined here, so problematic.

The second assumption that we need to make in order to motivate the sentience problem is **uncertainty about other minds**. On this assumption (which is widely accepted), we are not always certain which individuals are sentient. This uncertainty has scientific as well as scientific as well as

---

philosophical sources. Scientifically, we are not always certain which individuals experience
pleasure and pain. For example, many people now accept that all vertebrates experience pain.\textsuperscript{7} But the category of invertebrates is much less clear. With respect to some species, like squid, many people think that the behavioral and physiological continuities with humans outweigh the discontinuities, and so they proceed on the assumption that these animals do experience pain. With respect to other species, like ants, many people think that the behavioral and physiological discontinuities with humans outweigh the continuities, and so they proceed on the assumption that these animals do \textit{not} experience pain (an inference that, we should note, is less well motivated). And with respect to still other species, like lobsters, many people think that the continuities and discontinuities are so well balanced that they have no idea what to think.

Moreover, philosophically, we are not always certain which individuals are conscious. Unlike your behavior and physiology, your private mental life is not publicly observable even in principle. Sure, I might \textit{perceive} you as having private, subjective experiences, and I might also \textit{infer} that you do by analogy with me. But this perception may be inaccurate, and this inference may be based on bad reasoning.\textsuperscript{8} Of course, we may think that the problem of other minds has a solution. But even if we do, it is not likely that this solution will ground \textit{certainty} about whether or not others are conscious in all cases. Instead, and at most, it will ground a high degree of confidence that, say, other humans are conscious, and then a decreasing degree of confidence that nonhumans are conscious depending on how behaviorally, physiologically, and evolutionarily continuous they are with us.

\textsuperscript{7} For discussion of evidence that fish experience pain, see Victoria Braithwaite, \textit{Do Fish Feel Pain?}, (Oxford: Oxford University Press, 2010).

\textsuperscript{8} For discussion of the problem of other minds, see Peter Carruthers, “The Problem of Other Minds,” in \textit{The Nature of the Mind: An Introduction} (Routledge, 2004), pp. 6-35.
If we combine these assumptions, the upshot is that we are not always certain which individuals have moral status. Depending on how persuaded we are by the problem of other minds, we might think that this uncertainty will last for a very long time, if not forever. Yet in the meantime, we still have to decide how to treat many individuals about whom we are uncertain. How many exactly? It is impossible to say for sure. But, to put things in perspective, the Encyclopedia Smithsonian estimates that, at any given time, “there are some 10 quintillion (1,000,000,000,000,000,000) individual insects alive.”\(^9\) It follows that, if there is even a one/10,000 chance that the average insect consciously experiences even one/10,000th the pleasure and pain that, say, the average dog does at any given time, then the expected total amount of pleasure and pain consciously experienced by insects at any given time is equal to that of 100 billion dogs. (And of course, we might think that these percentages should be higher.) It is therefore extremely important that, rather than wait for developments in cognitive ethology, comparative psychology, and philosophy of mind that may never come, we think as carefully as possible, right now, about how we morally ought to treat individuals in cases of uncertainty about whether or not they are sentient.

Since the question that we are asking here falls into the general category of the ethics of risk and uncertainty, we will proceed by considering the three main types of principles that one can accept in such cases – an incautionary principle, a precautionary principle, and an expected value principle – and by exploring the strengths and limitations of each principle as it applies to the sentience problem.

Before we do that, however, I think that it will be useful to provide a few clarifications and qualifications about my discussion in this paper. First, I will not be asking what we should believe about sentience and moral status in this paper. Instead, I will be asking what we should believe about sentience and moral status in this paper. Instead, I will be asking what we should
do in light of the beliefs that we have, i.e. how we should treat others in cases of uncertainty about whether or not they are sentient. For that reason, in what follows I will simply stipulate what our epistemic states are in particular cases, so that we can ask what follows for our moral obligations. (You may or may not endorse the epistemic states that I stipulate us as having, but that will not matter for our purposes here.) Of course, there are many important questions that we can ask about our epistemic states, including what we should believe about sentience, what we should do to improve our beliefs about sentience, and how we should evaluate individuals whose beliefs about sentience are false, irrational, or a product of negligence. But I will not be asking those questions here.

Second, I will not be asking how we should treat individuals all things considered in cases of uncertainty about sentience. Instead, I will be asking whether we should treat them as sentient, and therefore as having moral status, in cases of uncertainty about sentience. For that reason, in what follows I will focus primarily on simple cases that allow us to attend to the question at hand: whether or not we should treat the individual in question as mattering morally for their own sake, rather than merely for the sake of others. Of course, there are many other questions that we need to ask before we can know what to do all things considered in real cases, including: How much pleasure and pain does this individual experience? Are they rational or self-aware? Do they have projects and relationships? Are cluelessness and demandingness problems for a moral theory? And of course: Which moral theory is correct?

Part of what makes this topic so difficult is that all of these issues are directly relevant to our interpretation and evaluation of possible solutions to the sentience problem. So it would be a mistake to bracket these issues entirely, as that strategy would lead to a seriously incomplete analysis. At the same time, it would also be a mistake to fully engage in all of these issues, as
that would lead to a seriously bloated analysis. As a result, my aim in what follows will be to strike a balance between these extremes. As I have said, I will focus primarily on simple cases that allow us to attend to the question at hand. But I will also, throughout the paper, gesture at how issues such as degrees of sentience, cluelessness, demandingness, and indirect / esoteric morality will be relevant to our interpretation and evaluation of these principles. Moreover, in my discussion of the expected value principle in section 5, I will consider utilitarian as well as Kantian interpretations, in order to give you a sense of how our background moral assumptions are relevant here. Of course, since other perspectives will yield other interpretations, this analysis will not be exhaustive so much as illustrative of the relevant issues.

Finally, and relatedly, I am focusing on sentientism about moral status in this paper for the sake of simplicity and specificity. But this problem arises for other theories of moral status too, as well as in cases of normative uncertainty about moral status. For example, if you accept rationalism about moral status, then you will need to ask how to treat individuals in cases of uncertainty about whether or not they are rational.\(^\text{10}\) If you accept biocentrism about moral status, then you will need to ask how to treat individuals in cases of uncertainty about whether or not they are alive.\(^\text{11}\) And so on. Moreover, if you are uncertain about which theory of moral status is correct, then this problem arises for you at multiple levels: For every potentially morally relevant property \(p\), you will need to ask not only how to treat individuals in cases of uncertainty about

---

\(^{10}\) For an argument that a rationalist should accept a precautionary principle in cases of uncertainty about human rationality but not in cases of uncertainty about nonhuman rationality, see Peter Carruthers, *The Animals Issue: Moral Theory in Practice* (Cambridge: Cambridge University Press, 1992), Chapter 5.

\(^{11}\) For a discussion of the metaphysics of life, see Fred Feldman, *Confrontations with the Reaper* (Oxford: Oxford University Press, 1994).
whether or not they have $p$ (if, in fact, $p$ is morally relevant) but also how to treat them in cases of uncertainty about whether or not $p$ is morally relevant.\(^{12}\)

My sense is that the discussion will have a similar structure across cases: Whether we are talking about sentientism, rationalism, biocentrism, and so on, the incautionary, precautionary, and expected value principles will be the main candidate solutions, and these solutions will have similar strengths and limitations. However, I also think that the details of the discussion will differ from case to case, and therefore our all things considered conclusion might differ as well. For example, suppose that we know more about rationality than about sentience. In this case, we might regard cluelessness considerations as less pressing for rationalism than for sentientism. Similarly, suppose that fewer individuals are possibly rational than possibly sentient. In this case, we might regard demandingness considerations as less pressing for rationalism than for sentientism. And so on. As a result, I will not try to discuss all theories of moral status in this paper. But I do hope that my discussion about sentientism will serve as a blueprint for how other, similar discussions might go.

With these clarifications and qualifications in mind, I will now present and evaluate the incautionary principle, the precautionary principle, and the expected value principle.

---

\(^{12}\) Many philosophers think that talk of moral status is unhelpful. For example, Ben Sachs argues that, since many features of a situation are morally relevant, we should talk about morally relevant features directly, rather than talking about them indirectly by means of the concept ‘moral status’. But note that, while I am talking in terms of moral status, this problem arises whether or not you prefer to speak this way: As long as you think that (a) certain features of a situation are morally relevant and (b) we are not always certain whether or not these features obtain, you will face the kind of problem I am discussing here. In any case, for more on this issue, see Ben Sachs, “The Status of Moral Status,” *Pacific Philosophical Quarterly* 92 (2011), pp. 87-104.
3. The incautionary principle

Start with the incautionary principle. This principle holds that, in cases of uncertainty about whether or not a particular individual is sentient, we are morally permitted to treat them as though they are not.

As far as I can tell, philosophers do not usually defend the incautionary principle in cases of risk and uncertainty, but I am starting with it anyway since many people seem to at least implicitly accept it in this area. So, what should we think about it? It faces at least two possible objections. The first objection is that the incautionary principle, as currently stated, is far too extreme. In particular, it implies that, if you are less than 100% certain that the problem of other minds has a solution (and therefore that other individuals have conscious experiences), then you are morally permitted to treat everyone in the world other than you as non-sentient. But I think that most of us would agree that this kind of moral solipsism is a non-starter. The mere fact that your friends and family members are only, say, 99.99% likely to be sentient, given your evidence, is not sufficient reason to treat them as though they are not.

A proponent of the incautionary principle might reply to this objection by pointing out that we can restrict the scope of the principle so that it applies in the case of, say, lobsters but not in the case of, say, our friends and family members. For example, perhaps we can say that the incautionary principle applies if and only if you are, say, less than 15% confident that a particular individual is sentient; otherwise one of the other principles that we will be considering here applies. Of course, it is a separate question whether or not this revision would be enough to rule out moral solipsism, as well as whether or not we could find a principled basis for this revision. But we can ignore those complications for the sake of argument here.
But even if we accept this reply, the incautionary principle still faces another, more important objection. Even if we restrict the scope of this principle in this kind of way, and even if we assume that our credences, or credence ranges, are precise enough to make this kind of restricted principle work, it still has implausible implications for our treatment of individuals who fall within its restricted scope. For example, suppose that you are only, say, 12% confident that a lobster is sentient, and therefore this lobster falls within the restricted scope of the incautionary principle. Suppose further that you feel inclined to boil this lobster alive – not even in order to eat him, but rather only for the simple pleasure of doing so. In this case, the incautionary principle implies that you have no moral reason at all not to act on this inclination, all else equal. However, I suspect that many of us think that this assessment of the situation is too simple: We think that, if there is a real chance that this lobster is sentient, then you need to take that possibility into account in your thinking about how to treat him. And if you do not – for example, if you boil this lobster alive for the simple pleasure of doing so despite believing that there is a 12% chance that he is phenomenally aware of every single moment of his torment – then your action is at least prima facie morally wrong. Moreover, what makes your action at least prima facie morally wrong is not only, as Kant claimed, that you might be conditioning yourself to be “harder in your dealings with” other human and nonhuman animals, but also, indeed primarily, that you might be harming this particular lobster here and now.\(^\text{13}\)

If this is right, then the lesson is: We need to find a solution to the sentience problem that makes our moral thinking sensitive to the possibility that a particular individual is sentient. The question is: How can we do that?

\(^{13}\) Immanuel Kant, “Duties to Animals and Spirits.”
4. The precautionary principle

This brings us to the \textbf{precautionary principle}. This principle holds that, in cases of uncertainty about whether or not a particular individual is sentient, we morally ought to treat them as sentient.

The precautionary principle, unlike the incautionary principle, is widely accepted. Many philosophers accept it in general, many animal rights advocates accept it in this area,\footnote{For an expression of the precautionary principle as a solution to the sentience problem in the animal advocacy community, see Gary Francione, “Sentience”: http://www.abolitionistapproach.com/sentience/#.Vg6p1J1Viko. (Thanks to [omitted] for this reference.)} and at least one philosopher, Alex Guerrero (2007) has defended a similar principle in this area, according to which you are blameworthy for killing individuals unnecessarily in cases of (subjective) uncertainty about whether or not they have moral status. So, what should we think of the precautionary principle? We can start by noting that this principle is much more plausible than the incautionary principle, for the simple reason that false negatives are worse than false positives in this area, i.e. accidentally treating a sentient individual as non-sentient is worse than accidentally treating a non-sentient individual as sentient. Therefore, a policy of erring on the side of caution in cases of uncertainty about sentience is better, overall, than a policy of erring on the side of incaution in such cases. For example, in the lobster case that we considered above, the precautionary principle implies, plausibly in my view, that you morally ought to treat the lobster as sentient, and therefore you morally ought to abstain from boiling him alive for the simple pleasure of doing so.

However, even if the precautionary principle is more plausible than the incautionary principle, it still faces two objections, which mirror the objections that the incautionary principle
faces. The first is that the precautionary principle, as currently stated, is implausibly extreme. For example, suppose that you are open to the possibility that panpsychism about consciousness is true, i.e. that everything in the world is conscious, and you are also open to the possibility that functionalism about pain is true, i.e. that the capacity to detect, and respond to, harmful stimuli is sufficient for the capacity to experience pain. Finally, suppose that, as a result of these expressions of epistemic humility, you are also open to the possibility that a very wide range of entities in the world – animals, plants, and things – are sentient. In this case, the precautionary principle implies that you morally ought to treat a very wide range of entities in the world – animals, plants, and things – as having full moral status. But one might object that this kind of moral animism is a non-starter as well. The mere fact that smart phones are, say, .01% likely to be sentient, given your evidence, is not sufficient reason to treat them as though they are.

A proponent of the precautionary principle has at least two possible replies to this objection (aside from insisting that they are not, in fact, open to the possibility that, say, smart phones are sentient). The first is to deny that this kind of moral animism is a non-starter. We have at least two possible reasons for taking this stance. First, we might claim that this objection is rooted in a concern about demandingness, and we might deny that demandingness is, in fact, a problem for a moral theory. On this view, if we should accept a precautionary principle in cases of uncertainty in general, then we should accept a precautionary principle in cases of uncertainty about sentience and moral status in particular, period – no matter how surprising or revisionary this result might first appear. (We will discuss this issue in more detail below.) Second, and relatedly, we might deny that this kind of moral animism is as demanding as it first appears. After all, as Peter Singer argues, equal consideration does not entail equal treatment. For instance, if we think that humans have a capacity to vote whereas nonhumans do not, then we
can extend equal consideration to all animals without extending the right to vote to all animals. Similarly, if we think that plants, if sentient, are still less sentient than we are, with dimmer pleasures and pains and fewer desires and interests, then we can extend equal consideration to all life without extending equal moral priority to all life. With that said, if we are uncertain about degrees of sentience as well, then we might not have this reply available to us, since we might think that we should accept a precautionary principle with respect to this issue too.

In any case, the proponent of the precautionary principle has a second reply available to them as well, which mirrors the reply that the proponent of the incautionary principle has available to them: They can restrict the scope of the principle so that it applies in the case of, say, lobsters but not in the case of, say, smart phones.\(^\text{15}\) For example, perhaps we can say that the precautionary principle applies if and only if you are, say, more than 5% confident that a particular individual is sentient; otherwise one of the other principles that we will be considering here applies. As before, it is a separate question whether or not this revision would be enough to rule out the relevant kind of moral animism, as well as whether or not we could find a principled basis for this revision (we will consider a Kantian attempt to find such a principled basis below). But we can ignore those complications for the sake of argument here.

But even if we accept one or both of these replies to the first objection, the precautionary principle still faces another, more important objection. Whether or not we restrict the scope of this principle, and even if we assume that our credences, or credence ranges, are precise enough to make this kind of restricted principle work, it still has implausible implications for our treatment of individuals who fall within its restricted scope. For example, suppose that you are, say, 12% confident that a lobster is sentient, and you are, say, 8% confident that a functionally identical robot lobster is sentient. (Why the difference? Because the real lobster is

\(^{15}\) Guerrero offers a version of this reply in 2007, p. 87.
physiologically and evolutionarily continuous with you whereas the robot lobster is not, and therefore you have at least a bit more reason to think that the real lobster is sentient than that the robot lobster is.) Finally, suppose that a house containing both lobsters is burning down, and you can save one and only one of them by pulling a lever that will direct water to their side of the house. Which lobster should you save? The precautionary principle implies that you are morally required to treat both lobsters as sentient, and therefore it is neutral about which lobster you should save all else equal. Yet I suspect that many of us think that this assessment of the situation is too simple: We think that, while both of these lobsters might be sentient, the real lobster is also a bit more likely to be sentient than the robot lobster is, given your evidence, and therefore you morally ought to break the tie in favor of saving the real lobster, all else equal.

This case illustrates a more general point, which is that there are costs involved with treating all possible sentience as actual sentience, without taking probability into account. That is, the more we expand the circle of moral concern, the more time, energy, and money we will have to divert away from humans, cats, dogs, cows, pigs, and so on and towards lobsters, insects, grass, trees, phones, laptops, and so on. Of course, we should not fall into the trap of thinking of morality as a zero sum game. Perhaps much more often than not, if we are thoughtful and strategic, we will be able to identify solutions to our social, political, and economic problems that do right by everyone rather than doing right by some at the expense of others. But there is no denying that we sometimes need to make hard choices. And here the numbers will end up mattering a lot.

For example, think again about the case of insects. As I said above, scientists estimate that there are currently about 10 quintillion insects alive. Thus, even if we assume that there is only a one/10,000 chance that the average insect experiences only one/10,000 the amount of
conscious pleasure and pain that the average dog does at any given time, the upshot is still that the total expected amount of conscious pleasure and pain experienced by insects at any given time equals that of 100 billion dogs. This is already a staggering figure. And if we remove probability of sentience from this equation, the figure becomes even more staggering: Specifically, it tells us that the total expected amount of conscious pleasure and pain experienced by insects at any given time equals that of $1e+15$ dogs, i.e. 1,000,000,000,000,000 dogs. (And of course, if we remove degrees of sentience from the equation as well, this figure becomes even more staggering still.) As a result, if we accept a precautionary principle for purposes of moral priority setting, we might find that, in comparison to the moral significance of invertebrates, the moral significance of vertebrates, including humans, becomes nothing more than a rounding error. Of course, we should be open to the possibility that this conclusion is correct. Still, given how much is at stake here, we might also want to examine this issue further, in part by asking whether or not this kind of all-or-nothing precautionary principle is perhaps too blunt an instrument to guide our thinking about moral priority setting, locally as well as globally.

If this is right, then the question that we now face is: Can we find a way to make our solution to the sentience problem sensitive to not only the possibility but also the (subjective) probability that different individuals are sentient? And if so, would this scalar solution to the sentience problem be better or worse than a (perhaps suitably restricted) all-or-nothing precautionary principle?
5. The expected value principle

This brings us, finally, to the **expected value principle**. This principle holds that, in cases of uncertainty about whether or not a particular individual is sentient, we morally ought to multiply our credence that they are by the amount of moral value they would have if they were, and to treat the product of this equation as the amount of moral value that they actually have. Of course, different moral theorists will flesh this principle out in different ways, and will face somewhat different challenges in the course of doing so. In what follows I will consider how utilitarians and Kantians might flesh it out, and I will also discuss the strengths and limitations of each approach. Along the way, I will consider how several other, related moral questions might affect our interpretation and evaluation of these principles, including questions about cluelessness, demandingness, and anthropocentrism.

Start with utilitarianism. Utilitarians are primarily concerned with how much conscious pleasure and pain we bring about. Therefore, a utilitarian will interpret the expected value principle as an **expected utility principle**, according to which, in cases of uncertainty about whether or not a particular individual is sentient, we morally ought to multiply our credence that they are by the amount of conscious pleasure and pain they would be experiencing if they were, and to treat the product of this equation as the expected utility of our treatment of them. If we can make this principle work, then it would have plausible results in many cases. For example, in the burning house case, it would imply that we should save the real lobster all else equal, which seems plausible. More generally, when we need to make hard choices about global priorities, this principle would imply that we should prioritize individuals about whose sentience we are more confident all else equal, which seems plausible as well.
There are two related objections that we might level against this expected utility principle, however. First, we might worry that, even if we set aside the problem of other minds, we are not capable of making precise, reliable comparative judgments in this area. This is true whether we make these judgments critically or intuitively. For example, think about what it would take to make them critically. We would have to calculate the probability that each individual is phenomenally conscious, and then calculate the probability that each individual experiences pleasure and pain. Then we would have to calculate the probability that each individual phenomenally consciously experiences pleasure and pain, based on our answers to these questions. I think we can all agree that it would be difficult for us to perform these calculations in any kind of precise and reliable way. Similarly, think about what it would take to make these predictions intuitively. We would have to make snap judgments about how likely each individual is to be sentient, and then rank them accordingly. But as before, we have very good reason to doubt that these snap judgments are reliable in many cases. For example, numerous studies have shown that our capacities for sympathy and empathy (as well as the judgments about sentience that they depend on) are sensitive to a variety of factors that we think, on reflection, are not particularly likely to be relevant, including big vs. small eyes, furry vs. slimy skin, four vs. six limbs, symmetrical vs. asymmetrical features, and so on. Thus, we might worry that our critical as well as intuitive judgments about sentience are not reliable.

---

16 For discussion of how difficult intersubjective comparisons of utility are even on the assumption that the entities in question are sentient, see Lori Gruen, “Experimenting with Animals, in Ethics and Animals (Cambridge: Cambridge University Press), pp. 105-129.

enough to be a useful guide. We might even think that, if we try to engage in these kinds of risk-benefit analysis in everyday life, then we will end up doing more harm than good overall.18

In many domains, this kind of worry about cluelessness is precisely what leads people to accept a precautionary principle instead of an expected value principle: They favor a precautionary principle not because they prefer a blunt instrument to a precise one, but rather because they think that, if you lack the time, energy, and information necessary to make precise, reliable judgments about expected utility, then a policy of erring on the side of caution is better, overall, than a policy of proceeding on the basis of what we can only assume is a deeply problematic, if not wholly arbitrary, judgment about expected utility.

A utilitarian might reply to this worry, however, by pointing out that, even if it would be a bad idea to try to make precise cardinal comparative judgments about sentience, it might not be a bad idea to try to make rough ordinal comparative judgments about sentience, at least in some cases. For example, in the burning house case, we do not need to assign a precise credence to the proposition that each lobster is sentient. Instead, all we have to do is break a tie between them. And, given that these lobsters are qualitatively identical except that the real lobster is physiologically and evolutionarily continuous with us whereas the robot lobster is not, we might think that a judgment to break the tie in favor of the real lobster is, at the very least, likely to be better than chance. Similarly, when we need to make hard choices about global priorities, we can grant that ideally we would be capable of making precise, cardinal judgments across species. But even if we cannot do this, perhaps we can at least make rough ordinal judgments across species, so that we can adopt a (very conservative, though still better than zero) discount rate across species in our decisions about resource allocation. As before, we might construct this kind of

18 For an argument that expected utility analysis concerning the distant future might do more harm than good overall, see Dale Jamieson, “Ethics, Public Policy, and Global Warming,” Science, Technology, and Human Values 17:2 (1992), pp. 139-53.
scalar, ordinal ranking with all due epistemic humility. But we might also think that this kind of scalar, ordinal ranking is, at the very least, likely to be better than no ranking at all.

Even if we accept this reply to the first objection, however, the utilitarian still faces a second objection, which is that, even if expertly applied, the expected utility principle is unacceptably anthropocentric. That is, if we follow this principle correctly, then we will end up systematically favoring individuals who are similar to us in certain respects, since, as we have seen, we have at least a bit more reason to think that individuals who are similar to us are sentient than individuals who are not, all else equal. Thus, for example, in a future world where we all co-exist with functionally identical robots, the expected utility principle would direct us to systematically favor real humans over robot humans, and it would direct robot humans to do the same. And we might worry that this kind of discrimination might not only be mistaken (since functionalism might be true) but might also result in the kinds of social conflicts that will do more harm than good overall.

In response to the first concern – that anthropocentrism is possibly wrong – utilitarians might simply accept that our epistemic standpoint is limited, and therefore the actions that we subjectively ought to perform, i.e. the actions that maximize expected utility, are not always the same as the actions that we objectively ought to perform, i.e. the actions that maximize actual utility. But in response to this second concern – that anthropocentrism, right or wrong, might do more harm than good overall – I think that utilitarians would, and should, be open to the possibility of accepting an indirect and/or esoteric decision procedure if need be. That is, they should say that, if, in fact, we can do more good by following a precautionary principle than by following an expected utility principle in certain cases, then we should follow a precautionary
principle rather in those cases. (This is what it means for our decision procedure to be indirect.\textsuperscript{19}) Similarly, they should say that, if, in fact, we can do more good by following the expected utility principle privately (i.e. secretly) than by following it publicly (i.e. openly) in certain cases, then we should follow it privately in these cases. (This is what it means for a decision procedure to be esoteric.\textsuperscript{20}) In this kind of way, the utilitarian might arrive at a hybrid decision procedure that directs them to use the expected utility principle in some situations and the precautionary principle in others, with varying levels of transparency.

Now consider Kantianism. Kantians are primarily concerned not about how much pleasure and pain we produce overall, but rather about whether or not we are treating subjects of moral concern as ends in themselves, with a “dignity beyond all price.”\textsuperscript{21} Therefore, Kantians will flesh out the expected value principle as an expected dignity principle, according to which, in cases of uncertainty about whether or not a particular individual is sentient, we morally ought to multiply our credence that they are by the absolute and incomparable moral value that they would have if they were, and to treat the product of this equation as the moral value that they actually have. As with the expected utility principle, the idea behind this principle is to take the possibility as well as probability of sentience into account in our thinking about how to treat others, but to hopefully do so in a characteristically Kantian way. The question is: Is it possible for a Kantian to strike the same kind of balance as a utilitarian here?\textsuperscript{22}

\textsuperscript{19} For discussion of indirect morality, see Richard Hare, \textit{Moral Thinking: Its Levels, Methods, and Point} (Oxford: Oxford University Press, 1982).
\textsuperscript{20} For discussion of esoteric morality, see Henry Sidgwick, \textit{The Methods of Ethics} (Hackett Publishing Company, 1981).
\textsuperscript{22} Some people might wonder why we are considering Kantianism at all here, since Kant was rationalist rather than sentientist, i.e. he believed that you have moral status if and only if you are rational, not if and only if you are sentient. However, philosophers such as Christine Korsgaard and Tom Regan have challenged this interpretation of Kantianism, and have developed
We can answer this question by considering, and replying to, two objections to the expected dignity principle. The first objection is that the expected dignity principle is not, in fact, a meaningful alternative to the precautionary principle, since both principles have the same implications in practice. Why might we think this? We might reason as follows: If moral status is absolute and incomparable, then it must also be infinite. (After all, how could a finite value capture the idea of dignity beyond all price?) And if moral status is infinite, then the expected dignity principle implies that, if we have any non-zero credence that a particular individual is sentient, then we morally ought to treat them as having full and equal moral status – since, if we multiply any non-zero number by infinity, the product is infinity. We might therefore think that the expected dignity principle collapses back into the precautionary principle in practice. For example, in the burning house case, we might think that it implies that we morally ought to treat both the real lobster and the robot lobster as having full and equal moral status, and therefore we morally ought to be neutral about whom to save all else equal. Similarly, when it comes to setting global priorities, we might think that the expected dignity principle implies that we morally ought to treat all species as having full and equal moral status, and therefore we morally ought to establish a discount rate of zero across species.

A Kantian might reply to this worry, however, by denying that moral status must be infinite in order to be absolute and incomparable. Yes, we can think of moral status as infinite if doing so helps us to respect the separateness of moral subjects by, e.g., refusing to aggregate sentimentist Kantianism as an alternative to the traditional rationalist interpretation. See, for example, Christine Korsgaard, “Interacting with Animals: A Kantian Account”, in Tom L. Beauchamp and R. G. Frey, eds., The Oxford Handbook of Animal Ethics (Oxford: Oxford University Press, 2011), pp. 91-118 and Tom Regan, The Case for Animal Rights (Berkeley: The University of California Press, 2004). My discussion in what follows will engage with this contemporary, sentimentist interpretation of Kantianism rather than with the traditional, rationalist interpretation.

23 The question what happens when we multiply zero by infinity is quite a bit more complex, but we can ignore that complexity for our purposes here.
across individuals. But ultimately we should not take this talk of infinity too seriously. As long as we respect the separateness of moral subjects in this kind of way, we can attach any arbitrary finite value to moral status for purposes of doing expected dignity analysis. And if this is right, then the expected dignity principle does not collapse into the precautionary principle after all. For example, in the burning house case – where, as a reminder, you can save one and only one lobster, and you are 12% confident that the real lobster is sentient and 8% confident that the robot lobster is sentient – the expected dignity principle, like the expected utility principle, implies that the real lobster has a higher expected moral value than the robot lobster does, all else equal. Similarly, when it comes to setting global priorities, the expected dignity principle, like the expected utility principle, implies that some species have a higher expected moral value than other species do, all else equal. As a result, Kantians, like utilitarians, can vindicate the intuitively plausible idea that we morally ought to save the real lobster over the robot lobster all else equal, as well as the intuitively plausible idea that we morally ought to save some species more than others all else equal – though of course, they would be quick to add that we cannot permissibly make use of these percentages for purposes of aggregation, as that would violate the kind of incomparability that Kantians take moral status to have.  

If you are not yet convinced that this interpretation of the expected dignity principle is compatible with the spirit of Kantian moral philosophy (since it still implies that we should treat some individuals as having more dignity than others, for certain purposes), it might help to consider an analogous case of risk and uncertainty. Suppose that two houses are burning down, and you have the power to prevent one and only one from burning down by pulling a lever that will direct water to that house. Suppose further that you have no idea if anyone is home in either house. All you know is that the houses are qualitatively identical, except that the house on the

---

24 Thanks to [omitted] for suggesting this reply to me.
left has lights on and music playing, whereas the house on the right has lights on but no music playing. What should you do? I am sure that different Kantians will answer this question differently. However, I also think that, if a Kantian says that you should prioritize the house on the left all else equal, that answer is not clearly incompatible with the spirit of Kantian moral philosophy. They would not, in providing this answer, be saying that people on the left, if home, have more dignity than people in the house on the right, if home. Instead, all they would be saying is that someone is more likely to be at home in the house on the left than in the house on the right, given your evidence, and therefore you are minimizing risk to people with full and equal moral status if you prioritize the house on the left, all else equal.

If this is correct, then we can think about the Kantian expected dignity principle in the same kind of way. When we use the expected dignity principle to show that you should prioritize the real lobster over the robot lobster in the burning house case all else equal, we are not saying that the real lobster, if sentient, has more dignity than the robot lobster, if sentient. Instead, we are saying that the real lobster is more likely to be sentient than the robot lobster is – i.e. that someone is more likely to “be at home” in the real lobster than in the robot lobster – given your evidence, and therefore you are minimizing risk to moral subjects with full and equal moral status if you prioritize the real lobster, all else equal. Of course, I am not taking a stand in this paper on whether Kantians should think about risk that way. Instead, what I am saying is that, if Kantians think about risk this way in general (as seems reasonable to me), then they can also think about it this way in this area in particular, in which case the expected dignity principle really does represent a meaningful alternative to the precautionary principle.

At this point, one might have a separate worry about the expected dignity principle, however. In particular, one might worry that, now that we have successfully distinguished the

---

25 Guerrero considers a similar case in 2007, p. 70.
expected dignity principle from the precautionary principle, the expected dignity principle inherits not only the strengths but also the weaknesses of the expected utility principle. First of all, as we have seen, we might worry that we are not capable of making precise, reliable comparative judgments in this area. As a result, if Kantianism requires us to make these comparative judgments in everyday life, then it will be difficult, if not impossible, to know if we are acting rightly according to Kantianism. We can call this the cluelessness problem. Second of all, as we have seen, we have at least a bit more reason to think that individuals who are similar to us in certain respects are sentient than individuals who are different from us in certain respects are, all else equal. As a result, if Kantianism requires us to rank individuals based on comparative judgments about sentience in certain cases, then it requires us to systematically favor individuals who are similar to us in certain respects over individuals who are different from us in certain respects, all else equal. We can call this the anthropocentrism problem.

Moreover, we can add a third problem to this pair, a problem which applies to the (Kantian) precautionary as well as expected dignity principles. In particular, both of these principles will require us to treat a staggering number of individuals in the world – at the very least, all human and nonhuman animals, including quintillions of invertebrates – as ends in themselves, with dignity beyond all price. It follows that Kantian morality is much more demanding than we might have thought. Granted, the expected dignity principle is not as demanding as the precautionary principle is, since the expected dignity principle permits us to favor vertebrates over invertebrates in certain cases all else equal (which, we might think, is more in line with our own desires and inclinations as rational agents). Still, the expected dignity principle does not allow us to make comparative judgments in many other cases; for example it would not allow us to treat insects merely as means to our ends (or to cause unnecessary harm to
insects in the course of pursuing our ends) simply on the grounds that we have more expected dignity than they do. As a result, we might find that, on the (Kantian) precautionary as well as the expected dignity principles, many standard human practices and traditions, including seemingly innocuous activities such as mowing the lawn or driving down the street, are morally impermissible. We can call this the demandingness problem.

Of course, these issues confront utilitarians too (as we discussed in relation to the first two issues above), but they are more acute for Kantians for a few reasons. First of all, one of the main strategies that utilitarians have for dealing with the anthropocentrism problem, which is to adopt an indirect and/or esoteric decision procedure if need be, is not available to Kantians who think that we cannot permissibly suppress our criterion of rightness in these kinds of ways. Second of all, utilitarians already accept that morality involves cluelessness and demandingness, since, after all, the world is both complex (hence cluelessness) and terrible (hence demandingness). In contrast, Kantians are reluctant to accept that morality might involve cluelessness and demandingness, since if I have no idea what to do, then I cannot act autonomously, and if I have to sacrifice all my projects and relationships for the sake of others, then I am treating myself merely as a means (among other reasons). The expected dignity principle – along with the precautionary principle, with respect to demandingness – therefore puts Kantians in an interesting dialectical position: Either they can adopt the utilitarian response to these problems, in which case they might have to change their theoretical position on publicity, cluelessness, and/or demandingness. Or they can maintain their theoretical position on these issues, in which case they will have to either (a) reject sentientist Kantianism as a result of these problems or (b) place restrictions on the expected dignity principle that allow it to avoid these issues altogether.
Which of these options are Kantians likely to take? The literature on risk and uncertainty suggests that many Kantians are likely to favor the latter response. In particular, this kind of issue usually plays out in that literature as follows. First, philosophers point out that risk and uncertainty pose a problem for Kantians, because risk and uncertainty are everywhere. Every time I drive a car, throw a ball, sled down a hill, or perform any number of other mundane actions, I accept at least a small amount of risk of harm or death to others. So, the challenge for Kantians is that, if we think of possible harm or death in the same kind of way that we think of certain harm or death, then the upshot will be that driving a car, throwing a ball, sledding down a hill, and generally speaking living our lives is morally wrong. Kantians usually then respond to this challenge as follows: They accept that cluelessness and demandingness are problems for a moral theory, and, as a result, they clarify (or revise, as the case may be) their moral theory in light of these concerns, by establishing a threshold of risk above which we must treat risk of harm or death as similar to certain harm or death, and below which we need not. Why? Because we all have a right to set ends for ourselves and take the means to them, we can all do only so much for others compatibly with that, and therefore we can bake cluelessness and demandingness restrictions directly into our criterion of rightness.

With respect to the anthropocentrism objection, in contrast, I imagine that many Kantians will simply bite the bullet and insist that morality is not about maximizing the good; instead, morality is about treating individuals as ends in themselves whether or not we thereby maximize

---

the good. As a result, many Kantians will say, if we morally ought to adopt a discount rate in certain situations, and if our adopting a discount rate in certain situations does more harm than good overall, then this harm is an unfortunate consequence of our activity but is not a reason to think that we have acted wrongly.

Of course, these moves are controversial, and what you think about them will determine, as well as depend, in part on what you think about the ethics of risk, uncertainty, cluelessness, demandingness, and the relationship between the good and the right more generally. I will not try to settle any of these issues here. My only aim in this section has been to indicate how utilitarians and Kantians might interpret the expected value principle, and what the strengths and limitations of this principle might be on these interpretations. If my discussion has been correct, then utilitarians and Kantians are in a similar position. The primary strength of the expected value principle, on either interpretation, is that it vindicates the plausible idea that we should take possibility as well as probability of sentience into account, at least in certain situations. However, the primary limitations of the expected value principle, on either interpretation, is that it commits us to much more cluelessness, demandingness, and anthropocentrism than we might have expected or hoped to encounter – unless, of course, we find a principled basis for restricting the principle in a way that avoids these implications. As I have indicated, utilitarians might restrict this principle as part of an indirect, esoteric decision procedure, and Kantians might restrict it to create space for agent-centered prerogatives. But even if they do, they will likely still find that this principle asks more of us, epistemically as well as practically, than we might have thought.
6. Conclusion

My aim in this paper has been to present and evaluate three possible solutions to the sentience problem: the incautionary principle, the precautionary principle, and the expected value principle (on utilitarian as well as Kantian interpretations). I have not attempted to present every possible solution to the sentience problem, nor have I attempted to show which, if any, of these principles is best all things considered. Nevertheless, I think that my discussion in this paper supports two provisional conclusions that I hope will guide future thinking about the sentience problem as well as about the problem of uncertainty about moral status more generally.

First, we should reject the incautionary principle. If there is a real chance that a particular individual is sentient (or has moral status more generally), then we have to take that possibility into account when thinking about how to treat them. It follows that, unless and until other, better solutions to the sentience problem become available, we should accept either the precautionary principle, the expected value principle, or both (for example, as part of a hybrid decision procedure).

Second, if we accept either the precautionary principle, the expected value principle, or both, then we should treat many individuals in the world as having at least partial moral status, including invertebrates, plants, robots, fetuses, PVS patients, and so on. Similar considerations might extend to certain collectives, such as insect colonies, as well. Of course, as I have emphasized throughout this paper, to say that we should treat these individuals (and collectives) as having at least partial moral status is not to say how we should treat them all things considered. For example, as Judith Jarvis Thomson famously argued, even if we assume that a

---

fetus has moral status, we might still think that abortion is morally permissible in many cases.\footnote{Judith Jarvis Thomson, “A Defense of Abortion,” *Philosophy and Public Affairs* 1:1 (1971), pp. 47–66.} Similarly, even if we assume that quintillions of insects (among many other individuals) have moral status, we might still think that killing them is morally permissible in many cases, for a wide range of familiar reasons. Still, if this conclusion holds, then we will have to raise moral questions about many standard, even mundane, human practices and traditions, ranging from how we clean our floors to how we get to work and back. And we will have to be open to the possibility that many of these practices and traditions will be revealed as at least prima facie morally problematic – and some might even be revealed as morally wrong all things considered.

The upshot, as I have argued throughout this paper, is that morality is likely to involve much more cluelessness and demandingness than we might have expected or hoped – unless, as I have said, we can find a principled basis for restricting our use of these principles in practice. Depending on which moral theory we accept, this cluelessness and demandingness will take different forms. For instance, since utilitarians accept aggregation, they might have to accept very revisionary conclusions about *global* issues. For example, they might have to accept that, even though the average human probably experiences much more pain than the average insect does, we should still prioritize insects over humans all else equal, since the total amount of expected pain experienced by quintillions of insects in the world easily trumps the total amount of expected pain experienced by seven billion humans. Conversely, since Kantians reject aggregation, they might have to accept very revisionary conclusions about *local* issues. For example, they might have to accept that, even though you probably experience much more pain than all the insects in your house combined, you are still required to treat each insect as an end in themselves, with a dignity beyond all price, and therefore you are not permitted to harm or kill
them merely as a means to your feeling comfortable at home, or as a byproduct of your doing so (though perhaps you would have moral reason to, say, prioritize yourself, or your cat or dog, over insects in certain rescue cases). Either way, then, we will likely find that a sincere commitment to acting on the precautionary and/or expected value principles will shape and constrain our choices in life in surprising and profound ways.\(^{30}\)

\(^{30}\) [Acknowledgements omitted.]