

The Moral Problem of Other Minds

Jeff Sebo (UNC-Chapel Hill)

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 directing our attention to here, which I will call the *moral problem of other minds* (since this problem is a moral analog of the conceptual and epistemological problems of

¹ David Foster Wallace, “Consider the Lobster,” *Gourmet Magazine* (August 2004).

² http://www.gourmet.com/magazine/2000s/2004/08/consider_the_lobster, p. 5.

other minds³) works as follows: Suppose that (1) sentience is necessary and sufficient for moral status, (2) we are not always certain if other beings are sentient, and, so, (3) we are not always certain if other beings have moral status. How should we treat other beings in such cases of uncertainty?

Wallace does not try to say how we should solve this problem in his essay. Instead, he argues that we have at least some reason to think that lobsters are sentient, and he concludes by expressing confusion about what to do all things considered. The philosophical literature is similar, though somewhat less candid. Many philosophers are aware of the moral problem of other minds, but few discuss it in detail, let alone try to solve it.⁴ Instead, most philosophers try to avoid this problem in one of two ways. First, they argue that particular beings (for example animals, plants, or fetuses) are sentient or are not-sentient.⁵ Second, they simply stipulate a solution to this problem (for example that we should err on the side of caution) without saying much about how this solution works in practice or why we should accept it.⁶

My aim in this paper is to present and evaluate three possible solutions to the moral problem of other minds. In particular, I will consider (1) an incautious principle that permits us to treat other beings as non-sentient in cases of uncertainty, (2) a precautionary principle that requires us to treat other beings as sentient in cases of uncertainty, and (3) an expected value principle that requires us to multiply our credence that other beings are sentient by the amount of sentience they would have if they were in cases of uncertainty. I will then argue for three general conclusions. First, the precautionary and expected value principles are more plausible than the incautious principle.

³ [Acknowledgement omitted.]

⁴ For an exception, see Alex Guerrero, “Don’t Know, Don’t Kill: Moral Ignorance, Culpability, and Caution,” *Philosophical Studies* 136:59–97, 2007. Guerrero argues that we should not kill individuals unnecessarily in cases of uncertainty about whether or not they have moral status. I think that this principle is compatible with both the precautionary and expected value principles that I will consider here.

⁵ See, for example, David DeGrazia, *Taking Animals Seriously: Mental Life and Moral Status* (Cambridge: Cambridge University Press, 1996).

⁶ See, for example, Tom Regan, *The Case for Animal Rights* (Berkeley: The University of California Press, 2004), p. 367.

Second, if we accept the precautionary principle or expected value principle, then we morally ought to treat many beings, including at least some invertebrates, as having at least partial moral status.

Third, if we morally ought to treat many beings, including at least some invertebrates, as having at least partial moral status, then morality involves more cluelessness and demandingness than we might have thought.

I will proceed as follows. In §2, I will present the background assumptions in moral philosophy and philosophy of mind that I will be making here, along with a few clarifications about the scope of my discussion. In §3-5, I will present and evaluate the incautionary, precautionary, and expected value principles, respectively. Finally, in §6, I will draw my conclusions and discuss the respects in which they might be morally revisionary.

2. Assumptions and Clarifications

I begin by presenting the assumptions in moral philosophy and philosophy of mind that I will be making here, along with a few clarifications about the scope of my discussion.

The first assumption that I will be making here is *sentientism about moral status*. On this assumption (which is plausible and widely accepted), a particular being has moral status, i.e. is a subject of moral concern, if and only if they are sentient, i.e. capable of phenomenally consciously experiencing pleasure or pain. What does it mean for you to be a subject of moral concern? It means that I can have moral duties to you (for example a duty not to harm you unnecessarily) for your own sake rather than merely for the sake of others. What does it mean for you to phenomenally consciously experience pleasure or pain? It means that when you experience pleasure or pain, there is

“something that it is like” for you to be having that experience.⁷ 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 moral problem of other minds so problematic.

The second assumption that I will be making here is *uncertainty about sentience*. On this assumption (which is plausible and widely accepted as well), we are not always certain if other beings are sentient. This uncertainty has scientific as well as philosophical sources. Scientifically, we are not always certain if other beings experience pleasure and pain. Granted, we can be relatively confident that other vertebrates experience pleasure and pain, since other vertebrates are relatively behaviorally, physiologically, and evolutionarily continuous with humans.⁸ But we cannot be as confident about invertebrates. With respect to some species, like squid, many people think that the continuities with humans outweigh the discontinuities, and so they proceed on the assumption that these animals do experience pleasure and pain. With respect to other species, like ants, many people think that the discontinuities with humans outweigh the continuities, and so they proceed on the assumption that these animals do *not* experience pleasure and pain (a conclusion that, we should note, is less warranted). With respect to still other species, like lobsters, many people have no idea what to think. And of course, similar questions arise, to a greater and lesser degree, for other kinds of beings as well, including plants, robots, fetuses, PVS patients, and even certain kinds of collectives.⁹

Moreover (and more importantly), philosophically, we are not always certain if other beings are phenomenally conscious. Unlike your behavior and physiology, your private mental life is not publicly observable even in principle. Sure, I might *perceive* you as having private, subjective experiences, and I might also *infer* that you do by analogy with me. But this perception may be

⁷ For discussion of the distinction between access consciousness and phenomenal consciousness, see Ned Block, “On a confusion about the function of consciousness,” *Behavioral and Brain Sciences* 18 (1995): 227–47.

⁸ For discussion of evidence that fish experience pain, see Victoria Braithwaite, *Do Fish Feel Pain?*, (Oxford: Oxford University Press, 2010).

⁹ For a discussion of the possibility of collective mentality, see Bryce Huebner, *Macrocognition: A Theory of Distributed Minds and Collective Intentionality* (Oxford: Oxford University Press, 2004).

inaccurate, and this inference may be based on bad reasoning, since, after all, we do not generally think that we are justified in drawing an inference about every member of a particular group (for example that all humans, or vertebrates, or animals are phenomenally conscious) based on our observation of a single member of that group (for example that I am phenomenally conscious).¹⁰ Of course, we may think that the problem of other minds has a solution, and therefore we may think that we can be relatively confident that at least some beings are phenomenally conscious. But even if we do, it is not likely that this solution will ground certainty about whether or not a particular being is phenomenally conscious in all cases. Instead, and at most, it will ground a high degree of confidence that, say, other humans are phenomenally conscious, and then a decreasing degree of confidence that other beings are phenomenally conscious depending on how relevantly similar they are to us.

If we combine these assumptions, then we are committed to *uncertainty about moral status*, according to which we are not always certain if other beings have moral status. Depending on how impressed 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 beings 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 (10,000,000,000,000,000) individual insects alive.”¹¹ It follows that, if there is even a one/1,000 chance that the average insect consciously experiences even one/1,000 the pleasure and pain that, say, the average dog does at any given time (which, I think, is a conservative assumption), then the expected total amount of pleasure and pain consciously experienced by insects at any given time is equal to that of 10 trillion (10,000,000,000,000) dogs. Moreover, many scientists and philosophers predict that artificial intelligence will soon rival human intelligence, which means that we will soon have to seriously consider the possibility that millions, billions, or even trillions of

¹⁰ For discussion of the problem of other minds, see Peter Carruthers, “The Problem of Other Minds,” in *The Nature of the Mind: An Introduction* (Routledge, 2004), pp. 6-35.

¹¹ http://www.si.edu/encyclopedia_si/nmnh/buginfo/bugnos.htm

artificial beings are sentient (and they will have to seriously consider the possibility that we are in turn). As a result, it is extremely important that, rather than wait for developments in science and philosophy that may never come, we think as carefully as possible, right now, about how we morally ought to treat others in cases of uncertainty about sentience and moral status. So that is what I will be attempting to do here.

Before I do that, however, I think that it will be useful to provide a few clarifications about my discussion in what follows. First, I will not be asking what we should *believe* about others given our *evidence* about them (since this more of an epistemic question than a practical question). Instead, I will be asking how we should *treat* others given our *beliefs* about them. Thus, for the sake of simplicity I will simply stipulate our level of confidence about whether or not a particular being is sentient, so that we can ask what follows for our moral duties to them. (You may or may not endorse the epistemic states that I will stipulate us as having, but that will not matter for our purposes here.) Of course, there are many related questions that we can ask about our epistemic states, including what they should be as well as how we should evaluate people whose beliefs 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 a particular being *all things considered* in cases of uncertainty about whether or not they are sentient (since that would require us to assess which theory of right action is true as well as canvass many other morally relevant features of the situation). Instead, I will be asking whether we should treat a particular being as sentient, and therefore as having moral status, in cases of uncertainty about sentience. Thus, for the sake of simplicity I will focus on cases that allow us to attend to the question at hand as much as possible, whether or not we should treat a particular being as mattering morally for their own sake, rather than merely for the sake of others. And, if and when we cannot answer this question without making at least some further moral assumptions, I will consider utilitarian and Kantian answers so that we can see how our

broader moral frameworks might affect our interpretation and evaluation of different solutions to the moral problem of other minds.

Finally, while I will be focusing on sentientism about moral status here, the moral problem of other minds arises, to greater or (usually) lesser degrees, for other theories of moral status as well. For example, if we accept rationalism about moral status, then we will need to ask how to treat particular beings in cases of uncertainty about whether or not they are rational. If we accept biocentrism about moral status, then we will need to ask how to treat particular beings in cases of uncertainty about whether or not they are alive. If we accept a relational theory of moral status, then we will need to ask how to treat particular beings in cases of uncertainty about whether or not they share the relevant kind of relation with us or others (for example contractual relationships or relationships of care). And so on. And of course, if we are uncertain about which theory of moral status to accept in the first place (for example if we are uncertain whether to accept sentientism or rationalism, or whether to accept this or that interpretation of sentientism), then we will have to ask this question at multiple levels, i.e. for every potentially morally relevant property or relation, we will have to ask both (1) how to act given our uncertainty about whether or not this property or relation is morally relevant and (2) how to act given our uncertainty about whether or not other beings have this property or share this relation, assuming that it is.

My sense is that the discussion will have a similar structure no matter what theory of moral status we accept: Whether we are talking about sentientism, rationalism, biocentrism, and so on, the incautious, precautionary, and expected value principles will be the main candidate solutions, and these solutions will have similar strengths and limitations. However, I also think some of the details of the discussion will depend on which theory of moral status we accept, and therefore our all things considered judgment about which solution to accept might depend on this issue as well. For example, if we think that we know less about the boundaries of sentience than we do about the boundaries of rationality, then we might think that *cluelessness* is more of a problem for a precautionary or

expected value principle in the context of sentientism than in the context of rationalism. Similarly, if we think that the boundaries of sentience are wider than the boundaries of rationality in the world, then we might think that *demandingness* is more of a problem for a precautionary or expected value principle in the context of sentientism than in the context of rationalism. As for uncertainty about which theory of moral status to accept in the first place: This kind of uncertainty shares some features with paradigmatic cases of descriptive uncertainty (such as uncertainty about whether or not a particular being is sentient) and other features with paradigmatic cases of moral uncertainty (such as uncertainty about whether or not a particular theory of right action is correct), and therefore it raises special questions that I will not be able to consider here. But while I will not be able to provide a complete discussion of descriptive and normative uncertainty about moral status here, I do hope that my discussion of the moral problem of other minds will be useful not only for sentientists but also for non-sentientists, since it can serve as a blueprint for how other, similar discussions might go.

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

3. The incautious principle

Start with *the incautious principle*. This principle holds that, in cases of uncertainty about whether or not a particular being is sentient, we are morally permitted to act as though they are not.

As far as I can tell, philosophers do not usually defend the incautious 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 incautious 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 incautious 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 incautious 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 incautious 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 incautious principle. Suppose further that you feel inclined to boil this lobster alive – not even in order to eat them, but rather only for the simple pleasure of doing so. In this case, the incautious 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 them. 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 they are phenomenally aware of every

single moment of this experience – 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.¹²

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 other beings are sentient. The question is: How can we do that?

4. The precautionary principle

This brings us to *the precautionary principle*. This principle holds that, in cases of uncertainty about whether or not a particular being is sentient, we are morally required to act as though they are.

The precautionary principle, unlike the incautious principle, is widely accepted among animal ethicists.¹³ So, what should we think of the precautionary principle? We can start by noting that precautionary principle is more plausible than the incautious principle in the lobster case that we considered above, since it implies, plausibly, that you have at least a prima facie duty to the lobster not to boil them alive for the simple pleasure of doing so. We can also note that at least part of what makes the precautionary principle more plausible than the incautious principle in this case, and in general, is that a false negative is worse than a false positive in this area, i.e. accidentally treating a sentient being as non-sentient is worse than the mistake of treating a non-sentient being as sentient all else equal.

¹² Immanuel Kant, “Duties to Animals and Spirits.”

¹³ See, for example, Gary Francione, “Sentience”:

<http://www.abolitionistapproach.com/sentience/#.Vg6p1J1Viko>. Relatedly, Alex Guerrero argues in (2007) that you are blameworthy for killing individuals unnecessarily in cases of (subjective) uncertainty about whether or not they have moral status. I think that this principle is plausible, but I also think that is neutral between the precautionary principle, which I consider in this section, and the expected value principle, which I consider in the next section.

Why should we accept that a false negative is worse than a false positive in this area? After all, false positives carry costs too. For example, insofar as we accidentally morally consider too many individuals, we run the risk of diverting scarce moral resources away from sentient beings who need them and towards non-sentient beings that do not. A proponent of the precautionary principle can answer this question in several ways. First, morality is not a zero sum game: Sometimes the act of helping or not harming someone replaces the act of helping or not harming someone else. But other times we can find creative ways to help, or not harm, a wider range of individuals with a static pool of resources. Second, insofar as morality *is* a zero sum game, the risk of morally considering too few individuals is greater than the risk of morally considering too many individuals all else equal, since the harm of receiving no moral consideration at all (which is the harm that moral subjects would experience in the former scenario) is greater than the harm of receiving too little moral consideration (which is the harm that moral subjects would experience in the latter scenario). Finally, if the risk of morally considering too many individuals is greater than the risk of morally considering too few individuals all things considered in a particular case (for example, if drawing the line at one place would yield a very large number of false positives and drawing the line at another place would yield only a very small number of false negatives in a particular case), then we can once again always build a restriction into the precautionary principle (or into our theory of right action) in theory or in practice in order to accommodate this concern.

However, even if the precautionary principle is more plausible than the incautious principle, it still faces several objections, which mirror objections that the incautious 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 phenomenal consciousness is true, i.e. that everything in the world is phenomenally 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 to the possibility that a very wide range of beings 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 beings in the world – animals, plants, and things – as having 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 currently, 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, current 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 cluelessness or demandingness, and we might deny that cluelessness and demandingness are, in fact, problems 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 involves as much cluelessness or demandingness as it first appears to. 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 experiences and fewer 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 degrees of sentience 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, current smart phones.¹⁴ 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 are 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 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. Assuming that you should save at least one of these lobsters, 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

¹⁴ Guerrero offers a version of this reply in 2007, p. 87.

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.

The third objection to the precautionary principle (which, as we will see, the expected value principle will face as well, though not as much as the precautionary principle does) is that the precautionary principle risks making morality unacceptably demanding. For example, think again about insects. As I said above, there are currently about 10 quintillion insects alive. For a utilitarian who accepts the precautionary principle, this means that if we are sufficiently confident that insects are sentient (for example if we set a risk threshold at 5% and think that insects are at least 5% likely to be sentient given our evidence), then we morally ought to treat 10 quintillion insects in the world as fully sentient, and, therefore, we morally ought to divert the vast majority of our moral resources to helping and not harming insects, assuming that we can do so effectively. Meanwhile, for a Kantian who accepts a precautionary principle, this means that if we are sufficiently confident that insects are sentient, then we have a moral duty to treat 10 quintillion insects in the world as ends in themselves, with a “dignity beyond all price,” and, therefore, we morally ought to abstain from many activities that we currently engage in, ranging from mowing the lawn to driving a car.¹⁵ Of course, we should be open to the possibility that this conclusion is correct. (Though, since utilitarians are more open to demandingness than Kantians, they might be more willing to accept this conclusion than Kantians, who might instead see this conclusion as a reason to either reject the precautionary principle or place further restrictions on it, as we will see below.) 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.

¹⁵ Immanuel Kant, *The Metaphysics of Morals*, ed. Mary Gregor (Cambridge: Cambridge University Press, 1996), 6:434-435.

If this is right, then the question that we now face is: Can we find a way to make our solution to the moral problem of other minds sensitive to not only the possibility but also the (subjective) probability that other beings 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 being is sentient, we are morally required to multiply our credence that they are by the amount of sentience they would have if they were, and to treat the product of this equation as the amount of sentience that they actually have. So, what should we think of the expected value principle? We can start by noting that, as with the precautionary principle, the expected value principle is more plausible than the incautious principle in the lobster case that we considered in section 3, since it implies that you have at least a prima facie moral duty to the lobster not to boil them alive for their own sake. We can also note that, as with the precautionary principle, at least part of what makes the expected value principle more plausible than the incautious principle in this case and in general is that a false negative is worse than a false positive – or, in the case of the expected value principle, a partial false positive – in this area.

How, then, does the expected value principle compare with the precautionary principle? It is hard to answer this question in the abstract, since the expected value principle will work very differently in the context of different moral theories. As a result, in what follows I will consider how it might work in the context of utilitarianism and Kantianism, and I will discuss the strengths and limitations of this principle in each case.

Start with utilitarianism. Utilitarians are primarily concerned with how much conscious pleasure and pain we bring about. Therefore, utilitarians will treat the expected value principle as an

expected utility principle, according to which, in cases of uncertainty about whether or not a particular being 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 then 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. As for interspecies comparisons of utility: If we think that insects are, say, 5% likely to be sentient given our evidence, then the total expected amount of insect pleasure and pain in the world on the expected utility principle is roughly 5% the total expected amount of insect pleasure and pain in the world on the precautionary principle. That seems plausible too.

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.¹⁶ 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

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.

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.¹⁷ Thus, we might worry that our critical as well as intuitive judgments about sentience are not reliable 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

A utilitarian might reply to this worry, however, by pointing out that, even if it would be a bad idea to try to make *cardinal* comparative judgments about sentience, it might not be a bad idea to try to make *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 decision to break the tie in favor of the real lobster is, at the very least, likely to be better than chance. Similarly, with respect to interspecies comparisons of utility: Even if we cannot make precise, cardinal judgments across species, 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 in our decisions about scarce resource allocation. As before, we might construct this kind of scalar, ordinal ranking with all due epistemic

¹⁷ For discussion of the many ways in which snap judgments about sentience can mislead, see Hal Herzog, “The Importance of Being Cute,” in *Some We Love, Some We Hate, Some We Eat* (Harper Collins, 2010), pp. 37-66. For a critique of anthropomorphism in cognitive ethology and comparative psychology, see Clive Wynne, “What are Animals? Why Anthropomorphism Is Still Not a Scientific Approach to Behavior,” *Comparative Cognition and Behavior Reviews* 2 (2007), pp. 125-135. For a defense of anthropomorphism in these fields, see John Fisher, “The Myth of Anthropomorphism,” in Colin Allen & D. Jamieson (eds.), *Readings in Animal Cognition* (Cambridge: MIT Press, 1996), pp. 3-16.

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.

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 this principle is unacceptably anthropocentric. That is, insofar as we follow this principle correctly, 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 to robot humans, and it would direct robot humans to systematically favor robot humans to real humans. And we might worry that this kind of discrimination might not only be mistaken (since functionalism might be true) but might also bring about the kinds of social conflicts that tend to cause a lot of harm.

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 would say that, if, in fact, we can do more good by following a precautionary principle than by following an expected utility principle in certain kinds of cases (for example in cases where expected utility judgments seem to be no better than chance, or where anthropocentrism would lead to harmful social conflict), then we should follow a precautionary principle in those kinds of cases.¹⁹ Similarly, they would say that, if, in fact,

¹⁹ For discussion of indirect morality, see Richard Hare, *Moral Thinking: Its Levels, Methods, and Point* (Oxford: Oxford University Press, 1982).

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 kinds of cases, then we should adopt a policy of following the expected utility privately in those kinds of cases.²⁰ In this kind of way, the utilitarian might arrive at a decision procedure that directs them to use the expected utility principle in some contexts 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.” Therefore, Kantians will treat the expected value principle as an *expected dignity principle*, according to which, in cases of uncertainty about whether or not a particular being 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?²¹

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 a Kantian precautionary principle, since both principles have the same implications in

²⁰ For discussion of esoteric morality, see Henry Sidgwick, *The Methods of Ethics* (Hackett Publishing Company, 1981).

²¹ Some people might wonder why we are considering Kantianism at all here, since Kant believed that you have moral status if and only if you are rational. However, philosophers such as Christine Korsgaard and Tom Regan have challenged this interpretation of Kantianism, and have developed sentientist Kantianism as an alternative to rationalist Kantianism. 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, sentientist interpretation of Kantianism.

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 we morally ought to treat everyone as having full and equal moral status in cases of uncertainty about sentience – since, if we multiply any non-zero number by infinity, the product is infinity. Thus, for example, in the burning house case we might think that the expected dignity principle 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, with respect to interspecies comparisons of dignity, 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 that we morally ought to establish a discount rate of zero across species all else equal.

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 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 is a meaningful alternative to a Kantian 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, all else equal. Similarly, with respect to interspecies comparisons of dignity, the expected dignity principle, like the expected utility principle, implies that some species have a higher expected moral value than other species, all else equal. If this is right, then 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 morality to involve.

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. Suppose that two houses are burning down, and you can 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 house on the left has lights on and music playing, whereas the house on the right has lights on but no music playing. What should you do?²² 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 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 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

²² Guerrero considers a similar case in 2007, p. 70.

“be at home” in the real lobster than in the robot lobster – given your evidence, and therefore you are minimizing risk to beings with full and equal moral status if you prioritize the real lobster, all else equal. To be clear, I am not taking a stand in this paper on whether Kantians should think about risk and uncertainty that way. Instead, what I am saying is that, if Kantians think about risk and uncertainty 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 a Kantian 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 expected dignity principle from a Kantian precautionary principle, the expected dignity principle inherits not only the strengths but also the limitations 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*.

We can also add a third problem to this pair, a problem which, as we have seen, applies to the precautionary as well as expected dignity principles. In particular, both of these principles 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 either way.

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 inclinations as rational agents). Still, the expected dignity principle does not allow us to make comparative judgments in many other cases; for example, if our credence that insects are sentient is high enough, the expected dignity principle does not permit 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 such as mowing the lawn or driving a car turn out to be morally wrong. We can call this the demandingness problem.

As we have seen, these issues confront utilitarians too, but they are more acute for Kantians for a few reasons. First of all, utilitarians already accept that morality involves excess amounts of 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 excess amounts of 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). Second of all, one of the main strategies that utilitarians have for dealing with cluelessness, demandingness, and anthropocentrism in practice, 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. The expected dignity and precautionary principles therefore put Kantians in an interesting dialectical position: Either they adopt the utilitarian response to these problems, in which case they might have to change their theoretical position on cluelessness, demandingness, and/or publicity. Or they reject the utilitarian response to these problems and maintain their theoretical position on these issues, in which case they have to either (a) reject

sentientism, (b) reject the expected dignity and precautionary principles, or (c) set limits on these principles that allow them to avoid these issues.

Which of these options are Kantians likely to take? The literature on risk and uncertainty suggests that many Kantians are likely to favor the third response. In particular, this kind of issue usually plays out in that literature as follows. First, philosophers point out that risk and uncertainty are everywhere.²³ Every time I drive a car, throw a ball, sled down a hill, or perform any number of other ordinary actions, I accept at least a small amount of risk of harm to others. So, the challenge for Kantians is that, if we think of possible harm in the same kind of way that we think of actual harm, then the upshot is 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 this result would reflect badly on a moral theory, so they clarify (or revise, as the case may be) their moral theory in order to avoid it, by establishing a threshold of risk above which we must treat possible harm as similar to actual harm, 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.²⁴

As for the anthropocentrism problem, I imagine that some Kantians will simply bite the bullet and say that, if we morally ought to adopt a discount rate in certain cases, and if our adopting a discount rate in certain cases 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. Other Kantians

²³ For an influential example of such an argument, see Robert Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1984).

²⁴ For an argument that we have a moral right to live our lives within certain limits, see Samuel Scheffler, *The Rejection of Consequentialism: A Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions* (Clarendon Press, 1984). And, for discussion of this issue as it relates to risk for deontologists, see Sven Ove Hansson, "Risk", *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2014/entries/risk/>.

might be reluctant to offer this response, however, since they might think that constructing a moral hierarchy based on how continuous other beings appear to be with us is incompatible with respecting the kind of dignity that other beings may well have. In that case Kantians will have to ask whether their concerns about the expected dignity people, for example that partial comparability appears to lead to anthropocentrism, are more acute than their concerns about the precautionary principle, for example that full incomparability leads to implausible results in many cases.

Of course, these issues are complex, and what you think about them will partly determine, as well as partly depend on, what you think about the ethics of 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 the expected value principle might function within utilitarianism and Kantianism, and what the strengths and limitations of this principle might be in each case relative to the precautionary principle. If my discussion has been correct, then utilitarians and Kantians are in a similar position. The primary strength of the expected value principle, for both theories, 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 limitation of the expected value principle, for both theories, is that it might be too difficult to apply in some situations, and it might be unacceptably anthropocentric in other situations. Meanwhile, both the precautionary and the expected value principles will turn out to be extremely demanding (though the precautionary principle faces this objection more acutely than the expected value principle does, given that the latter principle allows for epistemic discounting) – unless, of course, we find a principled way to avoid these implications in theory or in practice. As I have indicated, utilitarians might avoid these implications in practice by adopting an indirect, esoteric decision procedure, and Kantians might avoid them in theory by establishing a threshold for risk that creates 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

I think that my discussion in this paper supports three provisional conclusions that I hope will guide future thinking about the moral problem of other minds.

First, the precautionary and expected value principles are, relative to sentientism, much more plausible than 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 the precautionary principle, the expected value principle, or both, then we morally ought to treat many beings, including at least some invertebrates, as having at least partial moral status. Depending on our beliefs about sentience and threshold for risk, this list might also include insects, plants, robots, fetuses, patients in persistent vegetative states, certain kinds of collectives, and more. Of course, as I have emphasized throughout this paper, to say that we should treat these beings 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.²⁵ Similarly, even if we assume that quintillions of insects 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 human practices, and we will have to be open to the possibility that many of these practices will be turn out to be at least prima facie morally wrong – perhaps even morally wrong all things considered.

²⁵ Judith Jarvis Thomson, “A Defense of Abortion,” *Philosophy and Public Affairs* 1:1 (1971), pp. 47–66.

Third, if we morally ought to treat many beings, including at least some invertebrates, as having at least partial moral status, then morality involves more cluelessness and demandingness than we might have thought. Depending on which moral theory we accept, this cluelessness and demandingness will take different forms. For instance, since utilitarians accept aggregation, they will might have to accept 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 trumps the total amount of expected pain experienced by seven billion humans (though, as we have seen, utilitarians will face this issue more acutely insofar as they accept the precautionary rather than the expected value principle). In contrast, since Kantians reject aggregation, they might have to accept especially 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 them merely as a means to your feeling comfortable at home, or as a foreseeable consequence of your doing so, though perhaps you would have reason to, say, prioritize yourself, or your cat or dog, over insects in certain rescue cases (though, as we have seen, Kantians will be more justified in prioritizing vertebrates over invertebrates in rescue cases if they accept the expected value principle rather than the precautionary principle). Either way, then, we will likely find that a commitment to acting on the precautionary and/or expected value principles will shape and constrain our choices in life in surprising and profound ways.