In some circles, arguing, as Nicholas Agar does, that all life has intrinsic moral worth—that not only pigs, porpoises and protozoa are loci of moral value, but also species, rain forests, and ecosystems—will be considered a desirable and even noble project. In other circles, someone speaking with a straight face about the "moral goods of the bacterium or the fungus" is more likely to meet with rolling eyes. Members of both circles will find Agar's book interesting. Supporters of radical environmental ethics will, naturally, want to see an intelligent attempt to get some of their favored ideas up and flying. And those who find simply preposterous the idea of a patch of moss having inherent moral value also have much to learn here. Even those whose initial reaction is to judge the positive thesis quite bizarre should find the approach of this book well-informed, sober, and stimulating, for it is always of interest to see a novel argument for an unusual philosophical view.

Any such radical bioethical view of course raises a cluster of practical questions. How are we to weigh human interests against the interests of animals and plants? We're all familiar with the moral case for vegetarianism, but how are we to deal with the rights of the soybean plant or the humble cabbage? Will we even be permitted to walk to the kitchen for a bowl of muesli, if in doing so we crush underfoot a thousand morally valuable micro-organisms? Agar addresses such issues, arguing for a view that is metaethically radical in its extension of moral predicates to all living matter (including collectives of living things), but purportedly less extravagant in its normative output. He actually suggests little in the way of concrete prescriptions, but assures us that the upshot of his theory will be an ethic that requires some effort on our parts without imposing exorbitant practical burdens. In this brief critical assessment I will put aside issues concerning the normative output of Agar's project (about which much might be said), concentrating instead exclusively on the core of his metaethical argument. As will become apparent, I don't think the argument works.

The key to Agar's case is a notion of a "plausible naturalization" of a contested term. Consider a term like "function". Can we give a monolithic, all-encompassing naturalization of the notion, such that all and only the things to which we are firmly inclined to assign functions come out as having them? Agar thinks not. A better way of thinking of the situation, he suggests, is to consider a variety of naturalizations, each best suited to its own explanatory role in its own discipline. The biologist uses one naturalization of "function", the medical doctor another, the mechanic yet another; and these different naturalizations of the same term co-exist quite happily. A token object may be properly assigned a function relative to one domain, while being properly denied that function relative to another domain. Some usages may be closer to the vernacular usage than others, but that is not the test of adequacy—that test is, rather, whether they earn their keep within their respective domains.

The second step of the argument is to insist that if we're dealing with a term that in

vernacular usage has "value-endowing" properties ("function" is unlikely to be such a term), then at least some of that value will continue to attend the term in its various plausible naturalizations. The term "intelligence", for example, Agar takes to be a "morally loaded" term when used in ordinary contexts. It is also a contested term that admits of a variety of plausible naturalizations, each suited to a specific domain. The value that (putatively) is assigned to entities described in ordinary contexts as "intelligent" will be distributed, in differing degrees, among the entities described as "intelligent" in other specialist domains.

Agar's claim is not that any specialist use of a term from ordinary language must bring some of the vernacular connotations with it. Consider, for example, the way that physicists have co-opted ordinary terms like "flavor" and "color" in order to denote attributes of sub-atomic particles. Something which we may consider a central platitude of the term "flavor" in its ordinary usage—say, "Having a mouthful of chocolate may interfere with a person's capacity to discriminate flavors"—obviously has nothing to do with the physicist's usage. However, this observation isn't relevant to Agar's argument, since nobody in their right mind would mistake the physicist's definition of "flavor" for an attempt to *naturalize* (in any sense) the everyday notion.

After a fair bit of peripheral discussion, we discover (in chapter 5) that the notion that is really going to do the work for Agar concerns the kind of *representation* of the environment that any organism must employ if it's to respond to changes in its surroundings. According to a popular philosophical view, representational states may be assigned even to very simple creatures, such as bacteria and protozoa, allowing their behavior to be described teleologically, in terms of their content-characterizable ends. This Agar offers as one "plausible naturalization" of the concept *life*. An entity like a growth of crystals or a river, though affected by its environment, does not have states "whose biofunction is to produce specific changes to, or movements of, it in response to particular states of the environment" (p. 92), and thus such entities will not count as living. Agar's metaethical argument is not that value is transferred from the vernacular use of "life" to new specialist uses, but rather that value is transferred from vernacular uses of terms like "goal frustration" and "preference satisfaction" to new applications of these phrases. There must be two steps to the argument. The first is to establish that "goal frustration" is a "value-endowing" phrase when used in everyday contexts. The second is to establish that at least some of this value will transfer into any new application of the phrase that results from a plausible naturalization. Both steps, I think, are dubious.

First, it is never made entirely clear what it takes for a term to be "value-endowing". Throughout his book, Agar seems to take it for granted that certain vernacular terms are "value-anchoring" (e.g., "intelligent"), seemingly unaware of, or unwilling to engage with, a respectable kind of moral skepticism that doubts even this. Some philosophers, for example, wonder whether there is any such thing as objective moral value at all. Agar, it seems, is simply not addressing such people. He's content to appeal to the widely held and well-entrenched intuition that *of course* humans and other intelligent life-forms are intrinsically and objectively morally valuable—and from that point of agreement he'll try to expand the extension of the morally valuable. My objection is not with taking moral intuition as a starting point in moral debate—appealing to widely held intuitions as a means of defeating a moral

skeptic is a well-trodden avenue, and moral realists are frequently fans of coherentist epistemology (often under the title "wide reflective equilibrium"). But the appeal to "entrenched moral commonsense" (p. 101) plays an uncomfortable role in Agar's argument, for if one takes as a starting point the intuition that a creature with complex psychological states (like a human) thereby has intrinsic moral worth—not presenting an *argument* to that affect, but just treating it as self-evident—then one cannot just conveniently ignore the equally strong and equally entrenched intuition that a small patch of moss (for example) does *not* have intrinsic moral worth. We cannot take a pick'n'mix attitude to moral intuitions.

That's a general worry, but I have a more specific concern about whether there even *is* a widely held intuition that the term "goal frustrating", when used in application to humans, is "value-anchoring". Is it really true that when we say (for example) "Al Gore had his goals frustrated", that this is a value-laden comment? Certainly in making the claim one would not be implying that this frustration was *wrong*. I suppose we might allow that it follows that it was "a wrong for Gore". And from there one might try to argue that what is wrong for Gore must at least be a *consideration* in our decisions affecting him. But this last step is not self-evident, and may be resisted. (Indeed, much moral theorizing over the last two thousand years may be understood as attempting to convince rational doubters that this step follows.) One might just as easily say that frustrating Hitler's genocidal goals was a *prima facie* wrong, and should thus should have figured as such in our consideration of how to treat him. But that, I submit, is neither how we do make decisions, nor how we should. Speaking in the vernacular, whether the frustration of a person's goals is good, bad, or neutral, depends entirely on the content of those goals. What this suggests is that the concept of "having a goal frustrated" *per se* has no particular evaluative connotation. If one wants to highlight the apparent connection from "ϕ frustrated X's goal" to "ϕ was a wrong for X" that's fine, but it is far from obvious that in asserting the latter one would be making a value judgment (let alone a *moral* judgment). The mere fact that it contains the word "wrong" won't suffice; if I assert "John believes that euthanasia is wrong" I don't thereby evaluate anything.

But suppose, for the sake of argument, that when we use the predicate "... had his/her goals frustrated" in everyday contexts we *are* always implying some kind of moral judgment, or at least raising a morally relevant consideration. A serious problem for Agar's theory is this: In the vernacular usage, we can presume, the predicate will be applied only to persons—paradigmatically, human persons. However, this may be a very relevant factor in the evaluative component of ordinary usage. Perhaps what is bad about persons having their goals frustrated is that they *feel* bad about it, that they have their *conscious* plans thwarted, that they *suffer*.[1] In other words (again): it's not goal frustration *per se* that we evaluate negatively, but goal frustration *in persons*. Since Agar's argument starts by acknowledging an evaluation that invariably attends the vernacular usage of a term, and since that usage will always involve persons (we can assume), then it's going to be extremely difficult to tell whether it is "goal frustration" or "goal frustration in persons" that is the "morally loaded" term. If it's the latter, then Agar's argument never gets off the ground.

---

[1] This is just a simple example for brevity; to make it plausible we might have to come up with something a bit more complicated.

How would we decide between the two options? Presumably, with a round of thought experiments. We'd need to imagine a creature that has goals but is not a person—say, a lobster—and ask: would frustrating that creature's goals be bad? No doubt Agar will say "Yes!", but the problem now is that this would be begging the question. We cannot assume that frustrating a lobster's goals is bad, as a way of showing that the term "goal frustration" (*per se*) has evaluative connotations, as a premise in an argument designed to prove that frustrating a lobster's goals (*inter alia*) is bad.

To make things clearer, here's an analogy. Consider the sentence "X intended to abandon her new-born offspring to fend for themselves." In any vernacular use, "X" must denote an intelligent, deliberating being, since only such a being will satisfy the everyday criteria for "intending" things. And, of course, we think that intelligent beings that form such intentions are morally criminal, so any vernacular use of the phrase has strong moral overtones. But under certain plausible naturalizations of "having an intention", we might allow that a turtle or an insect might "form an intention to abandon her new-born offspring to fend for themselves", and when we do speak of a turtle's intentions in this manner, the evaluative baggage drops away. Abandoning their offspring is simply what turtles do, and even if we find upsetting the scenes of tiny baby turtles being picked off by the gulls, we hardly *morally reproach* the mother, even to the smallest degree! What this shows is that no evaluation attaches to the notion of "intending to abandon one's new-born offspring to fend for themselves" *per se*, but rather to the thought of *an intelligent, deliberating being* forming this intention.

Let's turn now to the second step of the argument, assuming that the above problems were solved, and that a vernacular application of the predicate "... had his/her goals frustrated" implies moral consideration for the subject. The question is: Why should this value transfer through all plausible naturalizations? Agar is happy to allow that maybe *very little* value transfers, but refuses (unaccountably, in my view) to admit that perhaps *none* will. In his defense he offers an analogy (p. 97). Some foods contain so little dietary fiber that for all practical purposes humans can treat them as containing zero; nevertheless, beyond a shadow of a doubt these foods really do contain some fibre. But this, though true, hardly helps to answer the question. If you suspect that the piece of rock in front of you contains no gold whatsoever, the observation that some rocks contain minute traces of minerals doesn't prove that there must, after all, be *some* gold in your rock!

Not only is the dietary fiber analogy unhelpful, but it reveals something somewhat odd about Agar's general attitude to "moral value". In many passages he speaks almost as if it were a substance or tangible characteristic that all living creatures must have a least some of (humans have a lot, dogs somewhat less, bacteria a teeny weeny bit). But the basic question of what a value *is*, and what it is to "have" some, is never explored. To many people's way of thinking, "value" is primarily a verb—it's something that we *do*—and the objects of these actions may be said, derivatively, to "have value". One strand of Agar's argument privileges this, since he claims *we* have chosen to endow some of our terms with value connotations. But at the same time he wants to play the moral realist, for he is trying to convince us that items that we (most of us) are *not* inclined to morally value *are*, unbeknown to us, morally valuable.

It's difficult to see how one can have it both ways. If things are valuable only through our decisions to describe them with "value-endowing" terms, then there's absolutely nothing to stop us choosing to use those same terms *without* the value connotations, in certain specialist contexts, should we choose to. Suppose that there's a very well-entrenched convention that when someone use the word "mean" (an example of what is often called a "thick evaluative term") she expresses disapproving attitudes. There is nothing to prevent us developing an additional convention according to which when "mean" is used in a particular context, or regarding a particular kind of object, or in a particular tone of voice, no such disapproval is expressed. (In some circles, to describe something as "bad" is to praise it.) It would be pointless to insist that we shouldn't do this, and just false to claim that we cannot. But what is different about the "value-endowing" terms that lie at the heart of Agar's argument? Even if our linguistic conventions decree that a vernacular use of "intelligent" or "goal frustrating" is morally loaded, the claim that this value *must* transfer to the specialist contexts in which these terms are used—that it's somehow beyond our control, such that we are *forced* to value entities that we otherwise are not inclined to, whether we wish to or not—strikes me as a very problematic claim.

The final problem with Agar's metaethical argument that I will mention here concerns the fact that if it worked at all, it would produce a far too liberal domain of moral concern. We've seen that his argument revolves around plausible naturalizations of terms like "life", "goal-directedness" and "preference satisfaction". But there's no reason to think that these are the *only* value-endowing terms that admit of various naturalizations.

Consider the biologist's specialist use of "altruistic" and "selfish", as used in phrases like "kin altruism" and "reciprocal altruism".[2] Kin altruism is when an organism's behavior is performed at some cost to the individual, but which nevertheless enhances that organism's reproductive fitness by benefiting its kin. A fine example of this is when a bee stings in defence of its hive. Its life is forfeited, but the other members of the hive (all genetically linked) have their probability of survival increased. Biologists are happy to call this behavior "altruistic", though it is clear that in doing so they are using the word very differently from its everyday usage. Speaking in an ordinary context, altruistic behavior is that which is done with a certain other-regarding *motive*. (Ordinarily, human actions that benefit others, but which were *intended* by the agent to benefit only herself, are not called "altruistic".) Since we may safely assume that apiarian activity is not properly motivated at all, a stinging bee will certainly not count as altruistic$_{vernacular}$, though it may certainly be altruistic$_{ev. bio.}$.

It is also pretty clear that the vernacular use of "altruistic" has strong evaluative connotations. Altruistic actions merit praise and reward (*ceteris paribus*); selfish actions merit disapprobation and chastisement (*ceteris paribus*). And this is where

---

[2] Actually, the naturalization of "selfish" that is discussed directly by Agar is Richard Dawkins' well-known application of the term to genes. But this strikes me as an unhappy choice of example, since Dawkins has insisted on several occasions that his use of "selfish" in this context is nothing other than a suggestive *metaphor*, and not like a scientist's revision, expansion, or naturalization of an extant term. But this doesn't undermine Agar's general point, and the co-opting of "altruism" by biologists like R.L. Trivers and W.D. Hamilton seems a better example.

problems arise in Agar's account, for if his logic is followed we appear to get to the conclusion that the stinging bee to some small degree merits reward (*ceteris paribus*), while a feeding bee, in acting selfishly$_{ev.\ bio.}$, to some small degree merits castigation. Only in Disney movies does such gross anthropomorphism have a place; it is not a desirable philosophical or ethical result.

And nor is it one that Agar embraces. Yet, after careful reading, I must admit to still being unsure how he avoids it while still pushing his argument home. First, he says that when value is transferred, it will not be "specific normative generalizations, but instead degrees of value carried by the ancestor notion" (p. 56). In other words (I think), perhaps the above analogy is unfair in that it countenances the transference of a very specific normative claim ("merits reward, *ceteris paribus*"), while Agar intends only very broad values ("is generally good"?) to pass. But quite why that should be so remains obscure to me; and, besides, it would be equally unacceptable to conclude that a bee's stinging behavior is in some general sense more "morally good" than its feeding behavior, as it would that the former merits reward.

Second, Agar claims that how much value transfers from the vernacular into specialist domains can be decided only after "thought experiments" have determined how far from ordinary usage the descendant notion is. So, for example, simple thought experiments can tell us that many actions that are altruistic$_{ev.\ bio.}$ are not altruistic$_{vernacular}$, and vice versa, and thus we can conclude that there is a great conceptual distance between the two. It appears that Agar would be happy to draw as a result that *very little* positive moral value attends the stinging bee's behavior. But it is less obvious why the result of thought experiments should not be that *no* positive moral value whatsoever does so. Indeed, the latter seems by far the preferable conclusion, as it does for my next example ...

Consider the way in which we use certain psychological terms in describing computer activity. (I don't mean fancy futuristic super-computers; but just the dusty old Mac sitting on my desk.) We say things like "The computer wants to print, but it doesn't know that the printer is connected to a different port, so it's waiting for you to tell it where the printer is". One might say that such talk is purely metaphorical, but it's also defensible to say that it's a "plausible naturalization" of some folk psychological terms in one very liberal manner that privileges systematic inputs and outputs. None of this would be to say that my desktop computer has beliefs and desires in the same way as a human has beliefs and desires, nor even in the same way as a dog or cat. But in each domain we may employ a somewhat different plausible naturalization, and find it of explanatory use to do so. I suggest that the very liberal understanding of "belief" and "desire" that allows me to ascribe such states to my computer can perfectly well "earn its keep" within its own domain.

Now consider the evaluative baggage that might go along with belief/desire talk when applied to humans in everyday contexts. We might uphold the principle "Generally speaking, one should give consideration to others' desires in making practical decisions". This is fine when speaking in a human-oriented domain. But that it might, *even to a small degree*, apply to the computer domain is no philosophical result to invite. When I decide to change printer ports on my computer, it's not a case of my important desires vastly outweighing my computer's trivial desires—rather, I should give *no consideration whatsoever* to my computer's desires. Indeed, for me

even to put my desires next to my computer's desires in practical decision-making is to make a grave mistake of muddling up importantly distinct linguistic contexts. Reflection on the possibility of a computer having its desires frustrated, far from forcing us to take the computer's desires into consideration, simply reveals that desire frustration is sometimes utterly value-neutral.

Agar is aware of such worries of liberalism, but as far as I can see he doesn't block them. For example, he argues that computers aren't going to count as having goals and representations according to his plausible naturalization, since his account specifically requires *biofunctions* and "self-directed" goals (p. 100). Quite so; according to the particular naturalization that Agar outlines in chapter 5, my computer has no representations, and thus no value. But his is just one plausible naturalization of "representation", and nothing is said to exclude others, such as the simple one I've outlined above. A naturalization that assigns desires to computers is perhaps somewhat further from the vernacular than the one that assigns them to bacteria and moss, and thus computers will probably have *less* moral value than living things, according to Agar. But the fact that they come out as having *any at all* is, I'm sure, a wholly unwelcome result.

The problem of liberalism is not just that the wrong things get assigned moral value, but things might get assigned the wrong kind of moral value. We've seen this earlier with the example of "altruism". In all Agar's discussion, the only kind of value that is mentioned is what we might call "moral subject" value: the interests of moss *demand consideration* in our decisions. But another kind of value—which we might call "moral agent" value—would also seem to attach to the vernacular uses of psychological terms. When we speak of a human's goals or desires, we might endorse the following principle: "Someone whose desires ignore everyone else's interests may be legitimately criticized" or "Someone who will pursue the satisfaction of his goals, even when it causes great distress to others, is bad". What's to stop these kind of values transferring through to dogs, moss, and bacteria? Clearly, any argument that leads to moral *agency* and *responsibility* being assigned to moss is a faulty one—moss just isn't the right sort of subject for such concepts. But nor, one might say, is it the right sort of subject for the concept *intrinsic moral value*.

Let me conclude. I have argued that the evaluative baggage that goes along with speaking of a human having her desires frustrated is unlikely to transfer to a computer having its desires frustrated, because in the former context the fact that it's *a human* (or, more generally, *a sentient being*) is a crucial element in that evaluation. Second, even if some "evaluative language" *does* transfer to the computer case (e.g., we might say "Ugh—you shouldn't have put that disk in; it didn't like that!") all this shows is that value terms can sometimes be used without the speaker *evaluating* anything at all. My point, of course, is that these considerations apply equally to Agar's arguments concerning living things. We should be quite prepared to accept that in a certain contexts it is permissible to speak of a patch of moss having goals, having preferences, and even its being possible to wrong the moss. These are all a long way from paradigmatic uses, but the terms in question are probably pliable enough to permit them. None of this, however, would show that the moss deserves *moral* consideration in our practical deliberations. On the contrary, it seems to me that the conspicuous conclusion is that it may be permissible to speak literally of an action

"wronging X" without this being in the least bit morally pertinent.

Agar's metaethical argument focuses closely on three things: a particular cluster of terms ("goal frustration", "preference satisfaction"), a particular plausible naturalization of these terms (the "bio-representational" account), and a particular kind of value that putatively attaches to the terms ("subject" value). Yet if the logic of value-transference works at all, then it should work with any of these variables altered. But we have seen that if we focus instead on the vernacular term "altruism" we get stinging bees being morally superior to feeding bees; if we focus instead on a different plausible naturalization of "representation"—one that doesn't require biology—we get morally valuable computers; and if we focus instead on the "agency" values that attach to vernacular terms we end up with computers *and* bees deserving reproach or praise. These are not conclusions we should accept; rather, we should reject any reasoning that leads to them.

There is a great deal in Agar's interesting book that I have not in the least touched on. In this brief critique I haven't tried to show that his conclusion is false, only that the argument he employs does not work. But it's probably pretty clear that I have little sympathy with radical environmentalist ethics. However, a rejection of the metaethical underpinnings hardly commits one to an indifference to the killing of wildlife, the destruction of ecosystems, or the extinction of species. Most of us, I'm sure, will agree that destroying the rain forest would be a bad idea; most of us, given the opportunity, will support measures to prevent that destruction occurring; perhaps we'll even go and chain ourselves to trees. But what is really added by insisting that our reason for holding such attitudes and for performing such actions is that the trees, the animals, indeed the rain forest as a whole, have *intrinsic moral value*? I must admit to having remained puzzled throughout Agar's book as to the background motivation of the radical biocentric metaethicist. What does one really hope to achieve in practical terms—as a motivational bulwark or as a persuasive stratagem— by earning the right for the use of the magical little word "moral" in discussion of our relationship with the environment?