1. Introduction
Suppose that pain is intrinsically morally undesirable, and that all agents have a non-instrumental moral reason to alleviate pain when possible. Now consider the following two cases:

Alice: Alice thinks very little about morality as such. However, for as long as she can remember, she has been deeply moved by the pain of others. Although Alice would not be able to articulate any justification or explanation for her attitudes towards pain, she is saddened by the thought that others are or might be in pain, and is motivated to alleviate their pain whenever possible. She gives a significant sum of money to the Guinea Worm Eradication Fund because she knows that by doing so she will be able to significantly reduce the amount of pain caused by this parasite.

Baxter: Like Alice, Baxter thinks very little about morality as such. Unlike Alice, he is not particularly interested in the alleviation of pain. However, he does care deeply about paying as little in taxes as possible. He gives a significant sum of money to the Guinea Worm Eradication Fund because he knows that by doing so he will be able to pay less in taxes.

In some ways, Alice and Baxter are similar. Neither is particularly concerned about morality as such. Both ultimately act in a way that is morally desirable. And yet we are likely to evaluate the character and conduct of these two agents very differently. Alice’s motivations should strike us as more virtuous than Baxter’s, and her action should strike us as reflecting something morally admirable about her in a way that Baxter’s does not. Why should this be so?

Several recently influential accounts of moral character offer a promising answer: Virtue consists at least partly in the possession of the correct attitudes towards that which is actually morally good and
that which is actually morally bad. These accounts differ in their details, but all evaluate agents on the basis of whether they desire, are motivated to pursue, or are otherwise for the actual good, as well as whether they abhor, are motivated to avoid, or are otherwise against the actual bad.1 The alleviation of pain is actually good, and, in caring deeply about pain, desiring to alleviate it, and being motivated to act accordingly, Alice shows that she is for the actual good and thereby virtuous. To save money on one’s taxes, in contrast, is morally neutral at best, and thus Baxter’s donation does not reflect any moral virtue.

But there is more than one way to conceptualize the actual good. Suppose that pain is identical to neural state P, and consider a third case:

**Charles**: Like Alice and Baxter, Charles thinks very little about morality as such. Unlike Alice and Baxter, Charles has been exposed to the concepts of neuroscience from a young age and has, for as long as he can remember, been deeply concerned with whether others are in neural state P. Although Charles does not know that neural state P is identical to or otherwise associated with pain or any other psychological state—and although he would not be able to articulate any justification or explanation for his attitudes towards it—he is saddened by the thought that others are or might be in neural state P, and is motivated to prevent or eliminate its occurrence whenever possible. He gives a significant sum of money to the Guinea Worm Eradication Fund because he knows that the guinea worm typically causes neural state P in its victims.

Since pain is identical to neural state P, Charles cares deeply about the alleviation of pain by virtue of caring deeply about the alleviation of neural state P. His action therefore reflects a deep concern for the actual good, and the aforementioned accounts of character may seem to imply that Charles displays just as much virtue as Alice. But most of us, I take it, will strongly resist this implication. It is difficult to accept that an agent’s concern for a particular neural state could be virtuous, unless he knows or believes that the neural state is associated with pain or some other psychological state.

The case of Charles suggests that the way in which an agent conceptualizes the actual good makes a difference to whether his attitudes towards it are virtuous. It seems that attitudes towards the actual good conceptualized in some ways are relevant to character, while attitudes towards the actual good conceptualized in other ways are not; call this the claim that certain conceptualizations of the actual good are or should be privileged. If we are to privilege certain conceptualizations of the actual good over others, then there must be some principled grounds on which to do so; call the need for such grounds the problem of conceptualization. To state this problem in the form of a question: What basis have we, aside from an appeal to intuitions, for claiming that Alice’s concern for pain conceptualized as pain is virtuous, while Charles’ concern for pain conceptualized as neural state P is not?

By way of foreshadowing: I defend no answer here, for the primary purpose of this paper is to show that this question is both difficult and important. That it is difficult is likely to be unclear upon first examination; in fact Nomy Arpaly and Timothy Schroeder (2014), in what to my knowledge is the only existing treatment of the problem of conceptualization, offer an apparent solution in the space of a few pages. The main goal of §2 is to show that their solution, which relies on the assumption that the correct normative theory identifies a particular conceptualization of the actual good as privileged, does not succeed. In §3, I discuss a number of other potential solutions and argue that each of these is also unsuccessful.

If we limit our discussion to toy cases like those of Alice and Charles, the importance of the problem may be similarly unclear. For in those cases it seems intuitively obvious how we should evaluate the agents in question; the difficulty there is merely one of providing a theoretical

---

justification for our judgments. But in §4, I argue that a number of other cases are affected by the same problem — cases in which our intuitions are likely to be unclear and in which the character of real-world agents is at issue. These include cases of ideologically motivated wrongdoers, who may display concern for the good conceptualized in one way but not in others, as well as the case of the psychopath, who may display indifference towards the bad conceptualized in one way but not in others. We might be tempted to avoid the problem altogether by abandoning the claim that attitudes towards the actual good matter to character, but in §5 I argue that a generalized version of the problem affects all accounts on which certain attitudes are intrinsically virtuous or vicious. I therefore conclude that we must either solve the problem of conceptualization or make the rather radical move to an account on which there are no such things as intrinsically virtuous or vicious attitudes.

2. Normative Theories and Normative Facts

I take Arpaly and Schroeder’s (2014) account to be broadly representative of those accounts of character which appeal to the actual moral good. On their account, virtue is a measure of the quality of an agent’s will; the quality of an agent’s will, in turn, depends on the extent to which he desires that which is actually good and that which is actually bad. There are several complications which can be dismissed after brief discussion here. One is that Arpaly and Schroeder treat good will and ill will as prior to and nominally distinct from virtue and vice. To have good will, on their account, is to desire that which is actually morally good or right; to have ill will is to desire that which is actually morally bad or wrong (162). They define virtue and vice in terms of good and ill will: To be virtuous is to possess significant good will without possessing ill will, and to be vicious is either to possess significant ill will or to lack good will in a significant way (202).

Their account actually requires that desires for the good or bad be both “intrinsic” and “correctly conceptualized” to qualify as good or ill will; I discuss these requirements shortly.

For our purposes, it is safe to ignore the nominal distinction between good will and virtue (and the corresponding distinction between ill will and vice). By way of explanation, it will be helpful to distinguish between two questions. First, what is it for an attitude to be virtuous, in the sense that its presence makes an agent’s character better than it would otherwise have been? Second, what is it for an agent to be a virtuous person? I read Arpaly and Schroeder’s account of good will to be addressed to the first question, and their account of what they call “virtue” to be addressed to the second. On their account, any instance of good will qualifies as virtuous in the sense that it makes better the character of the agent who possesses it. This is consistent with maintaining that the overall quality of an agent’s character must exceed a certain level — and thus that it must feature a certain preponderance of good will over ill — in order for that agent to be properly called a virtuous person. Because our primary interest in this paper is in whether particular attitudes are virtuous, I use “virtue” and “virtuous” in the sense that applies to attitudes, unless otherwise noted. I also treat “good will” and “ill will” as equivalent to “virtue” and “vice” in that sense. Thus I write that on Arpaly and Schroeder’s view, it is virtuous to desire the actual good and vicious to desire the actual bad.

A further complication is that Arpaly and Schroeder’s account of virtue is offered as a component of a larger project. Their ultimate goal is to provide an account of moral worth — that is, an account of the conditions under which agents are praiseworthy or blameworthy (or neither) for their actions. On their view, moral character and moral worth are very closely related: Agents are praiseworthy when and to the extent that their actions reflect virtue, and blameworthy when and to the extent that their actions reflect vice.

3. No doubt we are also interested in what is required for an agent to be a virtuous or vicious person, but this can be easily derived once we understand which attitudes are virtuous and vicious: On Arpaly and Schroeder’s view, the overall quality of an agent’s character is wholly a function of whether and to what extent he possesses virtuous attitudes as opposed to vicious ones.

4. More precisely, they are praiseworthy or blameworthy to the extent that their actions ‘manifest’ good will or ill will (or the absence of either) in a way
Our concern in this paper is moral character, rather than moral worth. And the following discussion does not require us to assume that there is any relationship between the two, much less to assume the very close relationship that Arpaly and Schroeder defend. If we do endorse something in the vicinity of their account of moral worth, then the problem of conceptualization should take on additional importance and perhaps additional urgency — by imperiling our ability to identify which attitudes are virtuous or vicious, it also imperils our ability to determine which agents are praiseworthy or blameworthy. The vignettes in this paper are constructed so that the agents in question not only possess certain attitudes, but also act on them; they are the kinds of cases in which the correct solution to the problem of conceptualization, if there is one, could determine whether or not an agent deserves to be praised or blamed.

But neither this feature of the vignettes nor this potential application of the problem of conceptualization should distract us from our main subject, which is moral virtue. In asking whether an agent like Alice or Charles has displayed virtue by acting, I do not take myself to be presupposing any claims whatsoever about moral worth; surely the question is an intelligible one even if it has no bearing on whether the agent in question is praiseworthy. In the discussion that follows — with the exception of a few instances in which I explicitly note otherwise — I leave all matters concerning moral worth to the side.

I take Arpaly and Schroeder’s account of character as a rhetorical starting point. But it is not necessary for us to agree with them that virtue consists entirely in desiring the actual good without desiring the actual bad. Nor is it necessary for us to agree that desires for the good and bad are the relevant attitudes, as opposed to some other attitudes or sets of attitudes which take the actual good and bad as their objects. The following discussion will be sufficiently motivated if we accept that at least part of virtue consists in possessing the appropriate “pro” attitudes towards the good and “anti” attitudes towards the bad. This modest claim, I hope, will seem plausible in light of Alice’s case — most of us should be willing to accept that she is a better agent than Baxter insofar as and because she cares about alleviating pain.

It will, however, be necessary for us to assume the truth of moral realism. In order for an agent’s attitudes towards the actual good to bear on her character, there must be an actual good. And the actual good, I assume here, is determined by normative facts which obtain independently of human conventions. The precise nature of the relationship between human moral theorizing and the underlying normative facts is the subject of considerable discussion later in this section; for now we merely require us to suppose that the attitudes of interest — the attitudes which we are trying to assess as either virtuous or non-virtuous — meet these additional requirements and are therefore part of the character of the agents who possess them.

---

5. In fact, this may be true even if the connection between moral worth and character is much weaker than Arpaly and Schroeder believe. Suppose that the expression of virtue is necessary but not sufficient for an agent to be praiseworthy. If we are required to satisfy an additional set of criteria, C, we can still construct cases in which the moral worth of an action depends on the answer to the problem of conceptualization: Simply imagine an agent who satisfies all the additional criteria in C, such that the determining factor in whether she is praiseworthy is whether the attitude she expresses is virtuous.

6. An anonymous reviewer suggests that, given the way the vignettes in this paper are framed, there is some risk of our intuitions about moral worth contaminating our reasoning about character. When we consider the case of Alice, for instance, we may first form the intuition that she is praiseworthy for having donated, and then conclude that she must have displayed virtue. Readers who are concerned about this possibility can eliminate it by changing all vignettes to feature agents whose (putative) virtue does not lead to action; they may simply imagine that Alice, for instance, cares deeply about alleviating pain but never has the opportunity to do so. My intuitions in response to these revised vignettes are the same as my intuitions in response to the originals, and I suspect that this will be common.

7. Note also that we are free to impose, if we wish, various conditions which an attitude needs to satisfy before it can properly be considered a part of an agent’s character — we may suppose, for instance, that such an attitude needs to have persisted for a certain period of time, or that it needs to be integrated to a certain degree with an agent’s other attitudes. These additional conditions have no effect on the problem of conceptualization, for they merely require us to suppose that the attitudes of interest — the attitudes which we are trying to assess as either virtuous or non-virtuous — meet these additional requirements and are therefore part of the character of the agents who possess them.
it is sufficient to assume that such underlying facts exist, and that it is in virtue of these facts that certain things are actually good.\(^8\)

As a final preliminary matter, there are several conventions employed in this paper which are worth explicitly noting. First, I frequently follow Arpaly and Schroeder in writing of an agent’s “desires” for the actual good. But I take no position here on whether desires specifically are the only relevant attitudes (or whether they are among the relevant attitudes); readers are free to substitute whichever “pro-\(^2\) attitudes they prefer in place of “desires”. Second, I write here of the actual “good”, but this is for simplicity’s sake and is not meant to imply that goodness is prior to or more basic than other moral properties, such as right; one can read “the actual good”, therefore, as shorthand for the “the actual good or the actual right”. Finally, I follow Arpaly and Schroeder\(^9\) in using capital letters to denote particular conceptualizations (164, fn. 10). Thus if I claim that an agent desires THE ALLEVIATION OF PAIN, I mean that the agent desires the alleviation of pain conceptualized as the alleviation of pain.

Arpaly and Schroeder’s solution to the problem of conceptualization is built into the formal articulation of their account of good will (which, again, I am treating as equivalent to moral virtue). On this articulation, good will consists in “intrinsic desire[s] for the right or good, correctly conceptualized...” (162).\(^{10}\) The correct conceptualization of the right or good is “to be determined by normative moral theory: the concepts employed in grasping the correct normative moral theory are the concepts through which one must intrinsically desire the right or good in order to have good will” (164).

Arpaly and Schroeder are apparently supposing that the correct normative theory presents the actual good using certain concepts. In the cases of Alice and Charles, for instance, the correct normative theory might present the good as THE ALLEVIATION OF PAIN. In order for an agent’s attitudes towards the actual good to be virtuous, they must employ the same concepts as the correct normative theory; thus, an agent would need to conceptualize pain as PAIN. Alice does desire that PAIN be alleviated, and so her desire for the good is virtuous. But Charles, though he desires that pain be alleviated, conceptualizes pain as NEURAL STATE P. Because the correct normative theory does not employ the concept of NEURAL STATE P — and because the identity of pain and neural state P cannot be “trivially deduced”\(^{11}\) — Charles’s attitudes towards the actual good are not virtuous.

The problem with Arpaly and Schroeder’s proposed solution, put briefly, is as follows. If we read “normative theory” literally and understand it to designate a kind of theory — a human artifact which attempts to describe the normative facts, and which uses particular concepts to do so — then there is no single correct normative theory. A normative theory which employs the concept of NEURAL STATE P might describe the underlying moral reality just as correctly as a normative theory which employs the concept of PAIN; it is therefore impossible to identify a particular conceptualization of the good by appealing to the correct normative theory. Alternatively, if we understand “correct normative theory” to designate the underlying normative facts which exist independently of human theorizing, then the correct normative theory does not use any particular concepts to identify the good. For

---

8. Note that when I refer to “normative” theories and “normative” facts, I mean the normative theories and facts which are relevant to morality and to moral reasons. Aesthetics, epistemology, and such other domains as might be taken to have a normative dimension are excluded from the present discussion.

9. They are themselves following Jerry Fodor (1998).

10. Italics added. In fact, this is Arpaly and Schroeder’s account of what they call “complete” good will; their distinction between complete and partial good will is irrelevant for our purposes. The requirement that desires for the good be “intrinsic” is merely meant to exclude cases of instrumental desires for the good, such as a desire to alleviate pain in order to make others like me.

11. If a particular conceptualization of the good is close enough to the “correct” one that an agent could trivially deduce that the two are conceptualizations of the same object, then Arpaly and Schroeder allow that desires conceptualized in this way could be virtuous (167). This does not factor into the following discussion, since the identity between pain and neural state P is not trivially deduced, and neither are the identities between the different conceptualizations which feature in later examples.
concepts are features of human attempts to describe reality, rather than features of that reality itself. On either understanding, the appeal to the correct normative theory fails to provide the needed justification for our judgment that Alice is virtuous but Charles is not.\footnote{12}

In the remainder of this section, I work through my objection to Arpaly and Schroeder’s proposed solution in more detail. First, a semantic point. I take it that on the most natural understanding of the term, a normative “theory” is just that—a theory which attempts to describe an underlying moral reality, and which is not to be conflated with the underlying reality itself. In this way, normative theories are like theories in other domains, such as the natural sciences. A theory in particle physics, for example, might describe the laws which govern the behavior of particles, but the theory is not identical to the laws themselves. Similarly, a normative theory attempts to describe the normative facts; I remain neutral here on whether these facts are best characterized as moral “laws”, or facts about what is valuable, or facts about the conditions under which agents have categorical reasons, etc. What is important is that the facts themselves are not identical to our attempts to describe them.\footnote{13}

Arpaly and Schroeder, of course, write of the “correct” normative theory, and one might wonder whether there is a meaningful distinction to be made between the normative facts and a perfectly correct description thereof. I contend that there is. For, on the most plausible understanding of “correct”, there can be multiple correct normative theories corresponding to a single set of underlying normative facts.

When we say that a normative theory is correct, I take it, what we mean is that it describes the underlying normative facts with perfect accuracy. And to describe these facts with perfect accuracy, I take it that a normative theory must accomplish both of the following. First, it must give a perfect accounting of where moral properties are instantiated in the world—it must tell us which actions are right or wrong, which states of affairs are good or bad, and so on. Second, it must give a perfect accounting of why the moral properties are instantiated where they are—it must tell us the property in virtue of which each right action is right, or the property in virtue of which each bad state of affairs is bad, and so on.

If these are the only two requirements for correctness, then it is possible for two or more distinct normative theories to be equally correct. Suppose that all and only those actions which alleviate pain are right, and that it is the fact that these actions alleviate pain which makes them right. Then, “Actions are right just in case and because they alleviate pain” is a correct normative theory: It tells us which actions are right, as well as the property in virtue of which they are right. But because pain is identical to neural state P, “Actions are right just in case and because they alleviate neural state P” is also a correct normative theory.

One might object that my second criterion for a correct normative theory—that the theory give an accounting of why the moral properties are instantiated where they are—is more difficult to satisfy than I have supposed here. For many of us will have the intuition that there is something about pain which makes the alleviation of pain right, and we might worry that the normative theory which commands us to alleviate neural state P does not adequately convey the reason that

\footnote{12}{In fairness to Arpaly and Schroeder, their remarks on conceptualization are primarily intended to address a rather different problem: They wish to require that agents desire that which is actually good or right, rather than the good or the right itself, in order to be virtuous. The appeal to normative theories is likely to be more successful for this purpose, since it would be a vacuous normative theory that simply commanded agents to “Act rightly” (See 164-165, 176-180). But they do seem to think that the appeal to the correct normative theory will also allow us to distinguish between agents like Alice and Charles (See 167). I am arguing here that their proposal fails to solve this secondary problem—as well as that what they apparently regard as a secondary problem is in fact a very important one.}

\footnote{13}{The analogy to the laws of physics is meant to be purely illustrative; I take it that this is the standard, naive view of the relationship between physical laws and physical theories, but I do not intend to endorse any position on scientific realism or the ontological status of physical laws here. If we are committed to moral realism, as I assume we are, then we should certainly distinguish between normative facts and normative theories, irrespective of our view of physics.}
alleviating pain is right. One might also object that, even if multiple
normative theories are equally “correct” in the sense described here,
there might nevertheless be reasons to view some of these correct the-
ories as better than others. I address both of these potential objections
in the next section, in the course of describing and dismissing several
alternative approaches to privileging a particular conceptualization of
the actual good. For now, I set these worries aside. If, as I have argued,
there are multiple normative theories which are equally correct, then
we cannot appeal to the single correct normative theory to identify the
privileged conceptualization of the good.

Although I have claimed that there are multiple correct normative
theories, I take it that only one complete set of normative facts obtains
in the actual world. So perhaps we should instead seek to identify the
privileged conceptualization of the good by appealing to these under-
lying normative facts, rather than to any particular theory about these
facts. The problem is that, unlike normative theories, the normative
facts do not describe the actual good by using any particular set of
concepts. This is because the normative facts do not describe the actual
good at all; they simply obtain. It is simply a fact that certain things are
good or bad, or that certain features make actions right or wrong, or
that we have categorical reasons under certain conditions, etc. If it is
a fact that pain is bad, then it is a fact that neural state \( P \) is bad, since
the two are identical and therefore share their properties. And there
seems to be nothing in the normative facts themselves to suggest that
one way of conceptualizing pain is correct at the expense of all other
ways.

It may be helpful to draw an analogy. WATER and \( \text{H}_2\text{O} \) are different
concepts, and yet water is identical to \( \text{H}_2\text{O} \). It is also a fact that water is
required for (most) fish to breathe. Suppose I were to ask the follow-
ing: Which conceptualization of water do the facts about fish biology
identify as the correct one? Do the biological facts pick out WATER
or \( \text{H}_2\text{O} \) as required for fish? These questions seem misguided, for the

facts of fish biology do not conceptualize water at all. The facts simply
dictate that a certain kind of liquid, which is both water and \( \text{H}_2\text{O} \), is
required for fish. I take it that the normative facts are similar. They dic-
tate that the alleviation of a certain state, which is both pain and neural
state \( P \), is right. And while we may understand these facts through the
use of various concepts, which concepts we use is a matter of indiffer-
ence from the perspective of the universe.

So I conclude that neither an appeal to the correct normative theory
nor an appeal to the underlying normative facts can justify our
judgment that Alice is virtuous while Charles is not. One might worry,
however, that this conclusion is premature. In the next section, there-
fore, I examine three possible attempts to solve the problem of con-
ceptualization: One which offers a putative basis for favoring some
correct normative theories over others, and two which appeal to the
underlying normative facts as grounds for privileging certain concep-
tualizations of the good.

3. Alternatives

(1) The Strawsonian Approach

I have argued that there can be multiple normative theories which
are equally correct. But one might point out that the sense in which
these theories are equally correct is a rather narrow one, and argue
that we have reason to view some correct theories as better than others.
Perhaps the most promising basis for such an argument is the appeal
to a broadly Strawsonian view of moral responsibility, on which the
appropriateness of praise or blame depends on the reactive attitudes
which we are prone to experience in response to certain conduct un-
der certain conditions. To be praiseworthy or blameworthy, on such a
view, is to have acted in such a way and under such conditions that the
community of human evaluators is disposed to react to one’s behavior
with certain kinds of positive or negative attitudes. An account of this
general kind is easily extended to describe moral character as well as

14. This may or may not have been what Arpaly and Schroeder originally in-
tended when appealing to the “correct normative theory”.

moral worth: To be virtuous or vicious is to have traits or attitudes that reliably evoke certain kinds of positive or negative reactions from the community of human evaluators.

A Strawsonian view of character would allow us to identify the privileged conceptualization of the good by working backwards. If we are reliably disposed to consider agents like Alice virtuous, but not agents like Charles — and this does indeed seem to be the case — then it must be an agent’s attitudes towards PAIN, rather than her attitudes towards NEURAL STATE P, that matter for the purposes of assessing moral character. Therefore it is the correct normative theory which employs the concept of PAIN that we should regard as best, and the conceptualization of pain as PAIN that we should privilege.

The problem with the Strawsonian approach as presented here is that it inverts the procedure proposed by Arpaly and Schroeder. They suggest that we start with the correct conceptualization of the good and use this to determine which agents are virtuous; the Strawsonian approach would have us start with our judgments about which agents are virtuous and use these to derive the correct conceptualization of the good. This inversion may not appear especially problematic when we are merely trying to make sense of Alice and Charles, as it seems intuitively clear which of these agents is virtuous and which is not. But, as I will argue in §4, the question of which conceptualization of the good is relevant is also likely to determine how we should evaluate other agents, such as psychopaths and ideologically motivated wrongdoers. Our intuitions about these agents are likely to be much less clear, and there seems to be no alternative but to turn to our best theoretical accounts of moral character for guidance. The approach described here apparently precludes the possibility of relying upon our theories in this way.

The basic Strawsonian premise, which affords a central role to our reactive attitudes, admits of further development. A sufficiently developed theory of reactive attitudes could enable a broadly Strawsonian account of character to avoid the problematic inversion described above. If such a theory provides an accounting of the conditions under which various reactive attitudes are appropriate, rather than merely describing the conditions under which we are in fact prone to possess these reactive attitudes, then it could render judgments even in those cases in which our actual reactions are muddled or vary across individuals.

There are many ways of filling in a Strawsonian framework. For our purposes, however, the details are significantly constrained: Our present interest is in those accounts on which the actual good plays a central role in determining which attitudes are virtuous, and for a Strawsonian account to be a live option for us, it must afford the actual good such a role. Whether there already exists a Strawsonian account which is fully suitable for our purposes is doubtful, but we may adapt one from Matt Talbert (2008), who holds that blame is appropriate in response to a wrongdoer who “[fails] to be appropriately sensitive to the moral reasons generated by other people’s needs and interests” (Talbert 2008: 530). We are interested in character rather than responsibility, and primarily interested in virtue rather than vice, so we may suppose that on our Strawsonian account, it is appropriate to judge virtuous an agent who is appropriately sensitive to the moral reasons generated by the needs and interests of others. If we further suppose that being appropriately sensitive to moral reasons is equivalent to desiring or otherwise being “for” that which is actually good, then our Strawsonian account will be very similar to Arpaly and Schroeder’s.16

But insofar as our Strawsonian account now resembles other accounts which appeal to the actual good, it no longer seems that it has

16 I think that this is a plausible, though not inevitable, interpretation of Talbert. Presumably the normative facts dictate that we actually have certain moral reasons with respect to others; e.g., that it is actually good or right to treat others in certain ways. And if to be sensitive to these reasons is to have attitudes that dispose one to act correctly in light of them, then sensitivity to moral reasons could be understood as an instance of desiring or otherwise having an appropriate pro-attitude towards that which is actually good. I have chosen Talbert’s account due to its potential to be interpreted as or adapted into an account like Arpaly and Schroeder’s, though the significance which Talbert places on “others” as a source of moral reasons complicates matters; see the end of §4 for a discussion of those accounts which evaluate agents based on their attitudes towards others.
any unique resources with which to solve the problem of conceptualization. The problem itself surely remains: It seems appropriate to judge Alice virtuous but inappropriate to judge the same of Charles, and our Strawsonian account must have some explanation for why this should be so. Both agents are appropriately sensitive to their moral reasons with respect to the pain of others; it is simply that one agent conceptualizes pain as PAIN and the other as NEURAL STATE P. And since our theory is no longer concerned with the conditions under which we actually make virtue judgments, it cannot appeal to those conditions as grounds to favor one of the correct normative theories over the others. Unless there is some other basis for favoring a particular correct normative theory — and I know of none which would be acceptable for our purposes — it seems that we must now turn to the normative facts to see if they ground any meaningful distinction between different conceptualizations of the good.

(2) The Appeal to Moral Understanding

If pain and neural state P are identical, then we cannot claim that the alleviation of the former is morally important while the alleviation of the latter is not — for, of course, the two are one and the same. But we might nevertheless claim that an agent’s attitudes towards ALLEVIATING PAIN should be privileged because the moral importance of pain can only be understood by those who conceptualize it as PAIN. The following seems intuitively plausible: In order for an agent’s attitudes towards the good to be virtuous, she must have some understanding of what’s good about the good. The goodness of alleviating pain, presumably, is due to the fact that pain feels a certain way. Thus an agent gets no credit for desiring to alleviate pain unless she understands that pain feels the way it does. And, one might think, to understand that pain feels the way it does is to conceptualize it as PAIN.

It is important to note that this strategy is, at its core, an appeal to the underlying normative facts: These underlying facts must somehow make true the claim that pain is bad because it feels a certain way. One way in which this claim might be made true by the underlying facts is if there exists some additional property — call it “dolorousness” — which is possessed by pain and by virtue of which pain is bad. With the aid of an additional assumption, it would then be easy to explain why an agent would need to conceptualize pain as PAIN in order for her attitudes towards it to be morally significant. Since pain and neural state P are identical, both have the additional property of dolorousness. But, we might assume, agents can only perceive or infer the fact that pain is dolorous when they conceptualize it as PAIN. Thus agents must conceptualize pain as PAIN in order to be aware of the feature which makes it morally important. This would justify our intuition that the desire to ALLEVIATE PAIN is virtuous (or, at the least, indicative of virtue, since it suggests a further desire to alleviate dolorousness), while the desire to ALLEVIATE NEURAL STATE P is not.

But this would also imply that it is dolorousness, rather than pain, that we have a non-instrumental moral reason to alleviate. Previously we assumed that pain was intrinsically bad, and that all agents had a non-instrumental moral reason to alleviate it. If we suppose that what is bad about pain is not the fact that it is pain but rather the fact that it is dolorous, then this accounting of the underlying facts cannot be correct. Pain is bad, but only insofar as it is dolorous; and what we actually have is an instrumental reason to alleviate pain, because to do so also alleviates dolorousness.

This set of normative facts is not necessarily implausible. We might understand dolorousness to be the phenomenally conscious experience of pain, while understanding “pain” to be its physical or otherwise
non-phenomenal correlates. Some hold that conscious experiences really are metaphysically distinct from their physical correlates, and it is not my concern to refute this possibility here. What is important to note is that the appeal to dolorousness represents a rejection of the identity principle which is required for the problem of conceptualization to arise in this case. It is not at all surprising that some conceptualizations of a given object are better than others at revealing the presence of a further, morally significant property. I may not be blameworthy for giving you what I conceptualize as \( \text{H}_2\text{SO}_4 \) while I would be blameworthy for giving you what I conceptualize as ACID. The reason for the difference, presumably, is that what is bad about giving you the acid is that it will harm you, and I know that acid conceptualized as ACID is harmful without knowing this about acid conceptualized as \( \text{H}_2\text{SO}_4 \). But, of course, the harmfulness of the acid is an additional property which is not identical to the property of being acid.

The interesting puzzle arises only when the good itself can be conceptualized in multiple ways. By identifying the good as the alleviation of dolorousness — rather than as the alleviation of pain/neural state \( P \) — we thus avoid the puzzle rather than genuinely solving it. And, as I will argue shortly, we should be searching for a genuine solution, for there are other cases in which the puzzle cannot be so easily avoided by denying the required identity principles.

Is there another way to make sense of our intuition that pain is bad only because it feels a certain way? Unless we understand “feeling a certain way” to describe some third property like dolorousness, it seems that we must understand it to refer simply to being painful. So pain is bad because it is painful; I read this as equivalent to the claim that pain is bad because it is pain. This fact gives us no basis to privilege PAIN over NEURAL STATE \( P \), for presumably any agent who grasps the concept of NEURAL STATE \( P \) knows that neural state \( P \) has the property of being neural state \( P \). And since being neural state \( P \) is identical to being painful, such an agent knows that neural state \( P \) is painful, and thus is aware of the property in virtue of which it is morally important.

(3) The Appeal to Counterfactuals or Conditionals
We are likely to have the intuition that there is nothing morally significant about a neural state like \( P \) except insofar as it is identical to some psychological state like pain. We have supposed, of course, that neural state \( P \) is identical to pain, so the normative facts which obtain in our own world provide no justification for this claim. But perhaps the underlying moral reality does support this intuition if we expand our conception of this reality to span across various possible worlds. For the following counterfactual claims may seem plausible: If neural state \( P \) were not identical to any psychological state, then there would be nothing morally important about it. But psychological states like pain and pleasure would still be morally important if they were not identical to neural states. That the moral importance of psychological states is robust across possible worlds might be taken as grounds for privileging the conceptualization of the good as ALLEVIATING PAIN over its conceptualization as ALLEVIATING NEURAL STATE \( P \).

I have been assuming that the identity between psychological and neural states is a metaphysically necessary one. If we maintain this assumption, then there are no metaphysically possible worlds in which pain and neural state \( P \) are distinct. Thus, a strategy which requires us to appeal to such possible worlds seems unacceptable. One might, however, suggest that there is another sense in which the identity of psychological and neural states is contingent — even if it is metaphysically necessary that pain is \( P \), it is not epistemically necessary. That is, an agent need not know that pain and \( P \) are identical, and thus it is, from

the perspective of that agent, epistemically possible that pain is not identical to neural state \( P \). This means that an agent could reasonably make the following assertion: “If it turns out that pain is not identical to neural state \( P \), then it is right to alleviate pain, but not right to alleviate neural state \( P \) (unless neural state \( P \) is morally important for some other reason).” Perhaps the fact that an agent could reasonably assert this conditional provides us a basis for identifying \( P \) as the relevant conceptualization.

However, this strategy fails. The problem is that the above conditional could reasonably be asserted only by an agent who has already adopted a normative theory (or other set of moral beliefs) which conceptualizes the good as ALLEVIATING \( P \) rather than ALLEVIATING NEURAL STATE \( P \). Consider another agent who is unaware of the identity between pain and \( P \), but who has adopted a normative theory (or other set of moral beliefs) which conceptualizes the good as ALLEVIATING NEURAL STATE \( P \). This agent could reasonably assert the following: “If it turns out that pain is not identical to neural state \( P \), then it is right to alleviate neural state \( P \), but not right to alleviate pain (unless pain is morally important for some other reason).” Thus, the conditionals which could reasonably be asserted by a given agent depend on the concepts employed by that agent’s normative theory (or beliefs), rather than on any feature of the underlying moral reality.

We could, of course, abandon the idea that the same conceptualization of the good is privileged for all agents. Perhaps different conceptualizations of the good are privileged for different agents, depending on each agent’s background beliefs about morality. Perhaps, for instance, the conceptualization of the good as ALLEVIATING \( P \) is privileged for an agent who believes that \( P \) is morally important; in contrast, the conceptualization of the good as ALLEVIATING NEURAL STATE \( P \) is privileged for an agent who believes that \( P \) is morally important.

The problem with this strategy is that it apparently presupposes that the agents in question have background moral beliefs. What of those agents who, like Alice and Charles, have given little thought to morality as such and have no moral beliefs about either \( P \) or NEURAL STATE \( P \)? Perhaps we should say that for such agents neither an attitude towards \( P \) nor one towards NEURAL STATE \( P \) can be virtuous, but this amounts to the view that an attitude towards the actual good cannot be virtuous unless the agent in question explicitly believes that the object in question is morally important. If we were inclined to describe Alice as virtuous (as I supposed that we were), then our intuitions provide us with some reason not to adopt this view. And given that our current task is to solve the problem of conceptualization from within the framework provided by accounts like Arpaly and Schroeder’s, the strategy described here does not seem acceptable for present purposes: It entails the rejection of a core feature of these accounts, which hold that agents with appropriate attitudes towards the actual good can be virtuous irrespective of their explicit moral beliefs or lack thereof. It remains open to us, of course, to reject the framework provided by such accounts, and the difficulties posed by the problem of conceptualization may be taken as a reason to do so. In the last section, I return to this question and argue that a version of the problem affects a much wider range of accounts of virtue — including those which require an agent to possess explicit moral beliefs.\(^{20}\)

\(^{20}\) An anonymous reviewer suggests that Julia Markovits’ (2010) account of moral worth may support a response of the kind described here. On Markovits’ account, agents are praiseworthy when the reasons which motivate them to act are the reasons which actually explain why their actions are right. Significantly, Markovits identifies the reasons which explain why actions are right with the reasons which provide evidence that actions are right. One might therefore suggest that whether a given conceptualization of the good is relevant depends on whether beliefs which employ that conceptualization constitute evidence for a given agent that his action is right; whether or not these beliefs constitute evidence will depend on which background moral beliefs the agent has and on the concepts which these beliefs employ. In light of the problem described here, however, it seems difficult to reconcile this reading of Markovits with her commitment to the claim that an agent need not possess explicit moral beliefs for his actions to have moral worth (e.g., 208). There is another reading of Markovits: We could interpret “evidence” not as information which justifies the belief that an action is right, in the sense of providing subjective evidence of rightness, but instead as information which warrants the belief that an action is right, in the sense of providing objective evidence of rightness. The fact that an action alleviates pain warrants the
What of those of us who reject the suggestion that psychological and neural states could be metaphysically identical? As noted earlier, I am not defending any psychophysical identity principles here but merely asking us to assume one in order to illustrate the problem of interest. But one might insist that it is not only false that pain is identical to any particular neural state but also metaphysically impossible for it to be so, in which case it might be inappropriate to assume a psychophysical identity principle even for illustrative purposes. And, of course, if pain and neural state P were not really identical to one another, it would be possible for one of the two to be uniquely identified as important by the underlying normative facts. Thus, at least in the particular cases of Alice and Charles, the problem of conceptualization would dissolve.

It is important to note that, even if we reject the possibility that psychological and neural states are identical, the problem of conceptualization does not disappear entirely. The cases of Alice and Charles are intended to illustrate the problem as clearly as possible, but the problem also arises in difficult and realistic cases, as I will discuss in the next section. These cases do not require us to assume any controversial psychophysical identity principles, but instead rely on the much less controversial assumption that persons are identical to beings with the personhood-conferring properties.

4. Realistic Cases

I introduce the following case elsewhere in the context of a discussion of psychopathy:

Clinic Bomber: George is deeply concerned with the well-being of persons, and is strongly motivated to save persons from being killed. Because he believes that fetuses are persons, he believes that he can save persons by preventing abortions. Accordingly, he places a small bomb in an abortion clinic and detonates it at a time when he knows the clinic will be unoccupied. The resulting damage to the facility, which is located in an area with limited access to abortion, forces it to close for several weeks and prevents a number of abortions which would otherwise have taken place.

George originally became convinced that fetuses were persons when he read a description of their biology. A fetus has a complete and unique human genome, and this, George thinks, endows it with personhood. This belief is false—a being actually requires certain psychological properties in order to be a person—but George came to have it by reasoning responsibly and without self-deception. Fetuses do not in fact have the required psychological properties, and so George’s action does not actually save any persons. (Clancy 2016: 764)

We may additionally suppose that George does not attribute any of the actual personhood-conferring properties to fetuses, and that, if fetuses were persons, George’s action would have been right. The question is how we should evaluate George’s character. There is some reason to think that George displays good will—or, for present purposes, that he displays virtue—since his action reflects the desire to save persons, which is actually good. On the other hand, there is reason to think that George does not display good will. After all, he takes persons to be beings with morally irrelevant genetic properties. Since it is a matter of moral indifference whether beings with these properties are saved, perhaps George’s desire to save them is itself morally neutral.
of how to evaluate George. To state the problem explicitly: George
cares about the actual good conceptualized as SAVING PERSONS,
but fails to care about the actual good conceptualized as SAVING BE-
INGS WITH X, Y, AND Z. Whether his attitudes are virtuous depends
on which of these conceptualizations of the actual good is relevant.

Alongside Clinic Bomber, I describe a second case meant to be
analogous:

Robbery: Newman is a psychopath. His general intelligence is average, and he has a fairly good understanding
of the minds of other humans. He understands that other
humans have feelings, a persistent sense of self, and all
of the other psychological properties that confer person-
hood. Newman does not, however, understand that other
humans have rights or moral standing, nor that they are
entitled to be treated in certain ways; therefore, he does
not understand that other humans are persons. Newman
robs a convenience store and intentionally kills the clerk
in order to avoid leaving behind any witnesses; he knows
that in so doing he kills a being with the personhood-
conferring psychological properties, but not that he kills
a person. (767)

I hold that psychopaths lack the concept of PERSONHOOD and
thus cannot know that their actions have any effect on PERSONS as
such. Yet psychopaths most likely do know that others have the prop-
erties which actually confer personhood. So a psychopath who de-
liberately harms another demonstrates that he is indifferent towards
HARMING BEINGS WITH X, Y, AND Z, but does not demonstrate
any attitude towards HARMING PERSONS.22 If harming persons is
actually bad, and if indifference towards the actual bad is vicious, then

21. In fairness to my past self, I did acknowledge the weakness of this argument
(see 765–766). There was less pressure in the rhetorical context of that paper
to distinguish between distinct properties and different conceptualizations of
the same property.

22. In this context, I assume that indifference is a kind of attitude or set of attitudes,
such that one cannot be properly indifferent towards HARMING PERSONS if
one lacks the concept of PERSONS. See Dana Nelkin (2015) on the distinc-
tion between indifference towards and insensitivity to moral considerations.
the correct evaluation of the psychopath depends on which of these conceptualizations of the actual bad is the relevant one.

I take it that our intuitions about these two cases are likely to be much less clear than our intuitions about Alice and Charles. Even if we as individuals happen to have clear intuitions, we should acknowledge that reactions to these cases are likely to vary widely across philosophers.23 Cases like these are especially worrisome because of the plausible connection between character and moral responsibility; recall that on Arpaly and Schroeder’s view, agents are praiseworthy for actions that reflect correct attitudes and blameworthy for actions that reflect incorrect ones. Thus, the question of whether clinic bombers display correct attitudes towards the good and the question of whether psychopaths display incorrect attitudes towards the bad may determine whether these agents are blameworthy as well as whether, and how severely, they may permissibly be punished.

Even if we find fault with the particular cases discussed here — if we, for instance, reject my claims about which concepts a psychopath does and does not possess — they should be sufficient to demonstrate that the problem of conceptualization can prevent us from reaching clear conclusions about realistic and important cases. These are precisely the kinds of cases in which we would like to be able to rely on our best theories of character to guide us. But theories which appeal to the actual good cannot guide us unless we have a procedure for determining which conceptualizations of the good to privilege. As I have argued, we do not yet have such a procedure.

In the final section, I argue that a generalized version of the problem of conceptualization affects all accounts on which certain attitudes are intrinsically virtuous. But it is worth pausing to note that the discussion of personhood in the present section already allows us to see how certain related accounts of character, which might be understood as appealing to the actual good in a somewhat different way, are also susceptible to the problem of conceptualization. Suppose that virtue consists not in the possession of certain attitudes towards the actual good but instead in the possession of certain attitudes, such as respect or care, towards others. Presumably, my respecting others (or caring about them, or possessing whatever other attitudes towards them might be appropriate) will entail that I also possess certain propositional attitudes, such as the desire that they not be in pain. But it might nevertheless turn out that respect for others does not itself reduce to any of these propositional attitudes or to any set thereof.25 And if it is the normative facts which determine that respect for others is an appropriate attitude, then such an account can still be interpreted as appealing to the actual good: Perhaps the reason that respect for others is appropriate is that others are, in fact, valuable, or that they are in some other sense deserving of respect.

The problem of conceptualization arises for such accounts because there are multiple ways of conceptualizing others and therefore multiple ways in which an agent might respect or care for them. I might conceptualize others as PERSONS or merely as BEINGS WITH X, Y, AND Z, and I therefore might respect or care about others under one of these descriptions without respecting or caring about them under the other.26 And insofar as the account in question is understood to appeal to the actual good — that is, insofar as it is the normative facts

23. The failure of philosophers to reach any consensus on the responsibility status of psychopaths has by this point become somewhat infamous; this is the primary subject of my 2016 paper.
24. While the case of George is more complex — the question in his case is whether his attitudes towards PERSONS are virtuous — I take it that the correct answer might still make a difference to whether it is appropriate to punish him. His expressed attitudes towards PERSONS might, for instance, ameliorate his blame for destroying property to a greater or lesser degree, or not at all.
25. See, e.g., Michael McKenna (2011) and David Shoemaker (2015) for accounts of responsibility which evaluate agents based on the attitudes they express towards others. See also Stephen Darwall (2002), who describes the attitude of care as having both direct and indirect objects, with a person’s welfare representing the direct object and the person representing the indirect object (1). It makes no difference to the present discussion whether the relevant attitude is respect or care, nor whether the attitude takes a person as a single direct object or as an indirect object along with other direct objects.
26. I make a similar point by way of response to McKenna and Shoemaker in my 2016 paper (770–771).
which identify others as the appropriate targets of certain attitudes — it seems that we have the same, limited set of resources for resolving the problem of conceptualization. We must look to the normative facts to identify which conceptualization of others is relevant to character — and if these facts cannot provide a basis for privileging PAIN over NEURAL STATE P, it is difficult to see how they could provide a basis for privileging either PERSONS or BEINGS WITH X, Y, AND Z.

5. Moving Forward

By way of conclusion, it is worth considering two other responses to the problem of conceptualization. One response is to conclude that all conceptualizations of the actual good are equally relevant. This would entail that Charles’s desire to alleviate NEURAL STATE P is virtuous. Although this implication is counterintuitive, perhaps it is one we should be willing to accept.

This seems unpromising, however. For if an agent’s attitudes towards the actual good and bad are relevant to virtue, then they are presumably relevant to vice as well. Suppose, as Arpaly and Schroeder hold, that indifference towards the actual bad is vicious,27 and further suppose that the infliction of pain is actually bad. We can construct cases in which an agent is indifferent to the infliction of NEURAL STATE P, and in which this indifference is reflected in her actions.

Suppose, for instance, that a doctor decides to give her non-verbal patient a certain kind of medication. She knows that the medication will cause NEURAL STATE P in her patient, but is wholly unconcerned, as she does not know that the medication will cause PAIN.28 This doctor is indifferent to that which is actually bad (the infliction of a state that is, in fact, identical to pain), and she displays her indifference by acting. We must therefore conclude that she acts viciously — unless we

27. Such indifference would, as I understand it, qualify as vicious insofar as it represents a lack of good will; assuming that the absence of the bad is good, to fail to desire the absence of the bad is to lack good will.

28. To rule out negligence, we may also suppose that she has no reason to suspect, and no way of knowing, that neural state P is identical to or associated with pain.

have some grounds to exclude certain conceptualizations of the actual bad as irrelevant. I take it that most of us will be much more resistant to attribute vice in cases like this than we are to attribute virtue in cases like Charles’. The problem is even more severe if we agree with Arpaly and Schroeder that agents are blameworthy when their vice is reflected in their actions, for we must then conclude that the doctor in this case deserves to be blamed and perhaps even punished — all for her failure to care about particular neural states.

Suppose that we are unwilling to accept the implication that agents like the doctor act viciously. A second response is to abandon entirely the idea that attitudes towards the actual good are relevant to character. If these attitudes make no difference to how we should evaluate agents, then there will be no pressure to privilege a particular conceptualization of the good; thus the puzzle would not need to be solved.

However, there is a generalized version of the problem of conceptualization which affects any account on which some attitudes are intrinsically virtuous or vicious. I follow Todd Calder (2007) in distinguishing intrinsic accounts of character from consequentialist accounts. On consequentialist accounts, the status of a given attitude or trait as virtuous or vicious is dependent on the consequences or expected consequences of its possession or exercise. If a certain attitude — say, the desire to alleviate pain — has or can be expected to have desirable consequences when possessed or exercised, then it is virtuous. But its status as virtuous is contingent. If circumstances were changed in the right way — so that, for instance, our attempts to alleviate pain almost always backfired by producing more pain — then the desire to alleviate pain would instead be vicious.29

On intrinsic accounts, the status of certain traits or attitudes as virtuous or vicious is intrinsic to the traits or attitudes themselves, rather than dependent upon the way in which these traits or attitudes interact with the external world. If it is intrinsically virtuous to desire to

29. See Linda Zagzebski (1996), Judith Jarvis Thomson (1997), and Julia Driver (2001) for defenses of consequentialist accounts of virtue; see also Ben Bradley (2005) for a description, but not a defense, of such a view.
alleviate pain, then no change in external circumstances could render this desire vicious.

I have implicitly assumed that an agent’s attitudes towards the actual good and bad, if they are virtuous or vicious, are intrinsically virtuous or vicious. This assumption corresponds to what I take to be a common intuition about such attitudes—that it is in an important sense good to desire such things as the alleviation of pain, and bad to desire such things as the infliction of pain, irrespective of the consequences of so desiring. Assuming that we share this intuition, our first instinct after abandoning accounts that appeal to the actual good may be to move to another kind of intrinsic account. That is, we might search for other attitudes which have a plausible claim to be intrinsically virtuous or vicious, rather than making the more radical move to an account on which the virtuous traits are identified by their consequences.

But any account on which certain attitudes are intrinsically virtuous or vicious is affected by a version of the problem of conceptualization. The reason for this is quite general: Any attitude must have an object, and any object can be conceptualized in multiple ways. Thus, in order to be useful for evaluating agents, any account which appeals to such attitudes must have something to say about which conceptualizations of the relevant object matter to character.30

For at least some promising alternatives, the question of precisely which conceptualizations are relevant is likely to be a difficult one. Take one alternative: Perhaps to care about the good or the right as such is virtuous.31 If this is correct, then agents are virtuous when they care about acting in ways that they believe to be good or right, irrespective of whether or not their imagined or intended actions have any features that would actually make them right.

Whatever the property of rightness turns out to be, there will presumably be multiple ways of conceptualizing it. Suppose, for instance, that for an action to be right is for it to be in accord with one’s categorical reasons—that is, the reasons which do not derive their force from any elements of one’s motivational set.32 That this is the correct account of rightness is not trivially deducible, and I take it that we can imagine agents who care about ACTING IN ACCORD WITH CATEGORICAL REASONS without caring about ACTING RIGHTLY, and vice-versa. Consider one such agent:

Estelle: Estelle is not particularly concerned about morality as such—she associates the moral domain with religious strictures which she rejects—and is therefore generally not motivated to perform actions simply because she believes them to be morally right. But she takes herself to be a rational person and she does care deeply about doing what she has most reason to do. Sometimes, Estelle feels that she has strong reasons to act which do not derive their force from any of her own goals or desires. She believes that the pain caused by the guinea worm provides her with such a reason to try and eradicate this parasite, so she gives a significant sum of money to the Guinea Worm Eradication fund.

To evaluate Estelle, we must ask whether she cares about acting rightly. She does; she cares that her actions have a property that is in dominant role, see, e.g., Ishtiyaque Haji’s (1998) account of moral worth.

30. An intrinsic account which does not appeal to attitudes of any kind would not face this problem. But I suspect that many of the traits which we might plausibly identify as virtuous or vicious ultimately derive their status as such from the virtuousness or viciousness of associated attitudes. Take, for instance, the tendency to excessive anger. Why should this trait be vicious? We might appeal to its behavioral consequences, but then our account would be a consequentialist account rather than an intrinsic one. The best alternative is to appeal to the fact that a tendency towards anger implies the presence of attitudes which are themselves intrinsically vicious.

31. For a view on which an agent’s attitudes towards the good as such play a
fact identical to the property of being right. But she conceptualizes this property as **BEING IN ACCORD WITH CATEGORICAL REASONS**, rather than conceptualizing it as **RIGHTNESS**. We will need some sort of procedure for determining which of these is the relevant conceptualization before we can reach any conclusions about Estelle.

Therefore, the move to an account which appeals to the good as such does not allow us to avoid the problem of conceptualization. It merely alters the problem: The difficulty now is in determining which conceptualization or conceptualizations of the good as such to privilege. **How** difficult this version of the problem will be depends on the broader framework of our new account of character and on why, specifically, it identifies certain attitudes as virtuous. On accounts which appeal to the actual good, our resources for solving the problem of conceptualization are limited: Since these accounts are theoretically committed to the claim that it is the normative facts that determine which attitudes are virtuous, we have only these facts (and, perhaps, the normative theories which describe them) to look to for guidance in identifying the relevant conceptualizations. Different sets of theoretical commitments are likely to provide different sets of resources.

I cannot canvas all possible intrinsic accounts of virtue here, of course. Nor can I explore the resources that might be available to these other accounts for solving the problem of conceptualization. It is possible that some such accounts provide for a satisfactory solution. ³³ It is also possible that there is some satisfactory solution to the problem even as it arises for accounts which appeal to the actual good. My aim here was to show that the problem is a difficult and important one, not that it cannot be solved. But the persistence of the problem does give us a reason to worry about accounts which appeal to the actual good, as well as a reason to be cautious when considering other intrinsic accounts. It also highlights an advantage of consequentialist accounts of character, and a reason, **ceteris paribus**, to prefer them. On such accounts, it is quite easy to determine which attitudes employing which concepts are virtuous. If the desire to **ALLEViate PAIN** — or to **ALLEViate NEURAL STATE P**, or to **SAVE PERSONS** — has or can be expected to have desirable consequences, it is virtuous. Otherwise, it is not.³⁴

### References


³³ An anonymous reviewer suggests one account which may have adequate resources available: Philip Stratton-Lake (2000) argues from a Kantian starting point that an action has moral worth only when the agent in question respects the universal law as such. Adapted to character, this suggests an account on which only an attitude towards the good conceptualized as **THE UNIVERSALIZABLE** is virtuous. Most likely Stratton-Lake’s account would require modification to solve the problem of conceptualization, for, as it stands, respect for the universal law is only *part* of what he thinks is required for an action to have moral worth — the action must also be motivated by the ‘concrete considerations’ which ‘[give] rise to particular obligations’ (62). In responding to these concrete considerations (e.g., the fact that a drowning person needs help), the agent responds to that which is actually good. Thus, Stratton-Lake’s account is a kind of hybrid which evaluates agents both on their attitudes towards (a particular conceptualization of) the good as such and on their attitudes towards the actual good. Insofar as the actual good plays a crucial role, the account still requires some explanation of which conceptualization of the actual good is relevant.

³⁴ Many thanks to two anonymous reviewers for their comments on this paper. This research was supported by a grant from China’s Fundamental Research Funds for the Central Universities program.