

Consequentialism, Procedural Ignorance, and Moral Theory[†]

Dale Dorsey

Department of Philosophy
University of Kansas
1445 Jayhawk Boulevard
Wescoe Hall, rm. 3090
Lawrence, KS 66045
ddorsey@ku.edu

DRAFT of 11/15/2011

Consider the moral theory called “objective act-consequentialism” (hereafter, “consequentialism”). This view holds that the moral quality of a given action ϕ in comparison to another action ψ is determined by the axiological quality of ϕ ’s actual consequences in comparison to the axiological quality of ψ ’s actual consequences. There are many reasons people have found to reject this sort of view. Some have claimed that it is too demanding.¹ Some have claimed that it disrupts a moral agent’s integrity.² Some suggest that it inappropriately denies that there are no constraints on the pursuit of the overall good.³

Other objections to consequentialism focus on consequentialism’s ill fit with the fact of human cognitive limitations.⁴ Traditionally, such objections focus on our general ignorance of the consequences of our actions, or our general inability to make precise expected utility calculations.⁵ However, a recent and penetrating objection argues that the problem of human cognitive limitations runs much deeper. Consequentialism seems to require us, at

[†]Special thanks are due to Doug Portmore, whose generous comments radically improved an earlier draft of this paper. I would also like to thank John Bricke, Richard DeGeorge, Julia Driver, Ben Eggleston, Erin Frykholm, and Scott Jenkins for helpful conversations.

¹Samuel Scheffler, “Morality’s Demands and their Limits” in *The Journal of Philosophy* 83 (1986).

²Bernard Williams, “A Critique of Utilitarianism” in Smart and Williams, *Utilitarianism: For and Against* (Cambridge: Cambridge University Press, 1974).

³Frances Kamm, “Non-consequentialism, the Person as an End-in-Itself, and the Significance of Status” in *Philosophy and Public Affairs* 21 (1992).

⁴Consider, for instance, James Lenman, “Consequentialism and Cluelessness” in *Philosophy and Public Affairs* 28 (2000).

⁵Cf. John Stuart Mill, *Utilitarianism* II.24.

virtually all times, to perform actions that, though strictly possible for us to perform, we will not perform given our limitations as cognitive agents. This is true even *if* we know for certain that the action in question will have tremendously good consequences, better than any other alternatives.⁶

In this paper I argue that this problem is not isolated; it applies not just to consequentialism, but to virtually every plausible moral theory. I argue that the root of this problem is to be found in what might be called a *theory of moral requirability*. I propose an independently plausible alternative to the traditional theory of moral requirability, and argue that this alternative successfully avoids commitment to a view that renders morally acceptable behavior out of reach for virtually all normal human beings.⁷

1. Cognitive Limitations and Moral Requirements

It's well-known that consequentialism has some sort of problem with actions one can, strictly speaking, perform, but actions that one won't perform given human cognitive limitations.⁸ Examples of this kind abound. Take, for instance:

Wilda: Wilda stands before a safe. Were she to open the safe, overwhelmingly good consequences would result. However, she has no idea what the combination of the safe is.

Despite the fact that Wilda doesn't know the combination of the safe, consequentialism would appear to require her to open the safe. After all, she *can* open the safe, i.e., by entering the right numbers in the right order. But it would seem odd to say that she behaved in a morally impermissible manner in failing to do so. After all, she doesn't know how.

Some consequentialists might feel comfortable simply biting this bullet. As Wiland notes, to say that Wilda is morally required to open her safe seems more or less innocuous, insofar as her case is pretty unusual, and given that consequentialism generally refuses to index an individual's moral

⁶Eric Wiland, "Monkeys, Typewriters, and Objective Consequentialism" in *Ratio* 18 (2005).

⁷This moniker is imperfect for a number of reasons; I use it here simply as a placeholder for the sort of objection I explore in the next section.

⁸See, for instance, Francis Howard-Snyder, "The Rejection of Objective Consequentialism" in *Utilitas* 9 (1997); Erik Carlson, "The Oughts and Cans of Objective Consequentialism" in *Utilitas* 11 (1999); Mozaffar Quizilbash, "The Rejection of Objective Consequentialism: A Comment" in *Utilitas* 11 (1999); Dale Miller, "Actual-Consequence Act Utilitarianism and the Best Possible Humans" in *Ratio* 16 (2003), 50; Wiland, *op. cit.*

requirements to his or her epistemic states.⁹ Somewhere along the line, however, this sort of bullet-biting starts to sound a bit goofy. Take another example:

Jeffrey: Jeffrey is spending a leisurely afternoon reading a titillating detective novel. He could, alternatively, sit down at his computer, type the cure for cancer, and send it to a top medical journal. The consequences of performing this action would be overwhelmingly good. It does not occur to Jeffrey to do this not because he believes that curing cancer would have suboptimal consequences, but rather because he has no clue what the cure for cancer is, or how to write it down.

Jeffrey, in a perfectly sensible sense of “can”, *can* widely disseminate the cure for cancer.¹⁰ All he needs to do is to sit down, type it, and disseminate it widely. Jeffrey can type. No one’s stopping him. Assuming such a cure exists, if we also make the relatively innocuous assumption that curing cancer would produce overwhelmingly good consequences,¹¹ it would appear that he morally ought to do so.¹² Not only this, but given the overwhelmingly good consequences of curing cancer, it would appear that simply continuing to read his detective novel is a *wildly* suboptimal action, an action that is *far* below the moral ideal.

But this is surely absurd. Though one might be willing to bite the bullet in Wilda’s case (though this doesn’t strike me as particularly plausible in any event), Jeffrey’s case illustrates that the problem at hand is not confined to the stranger than fiction. In fact, as Wiland puts it: “once we become literate creatures, the number of action-options usually available to us is astronomically high. . . The chance you ever do what in fact has the best consequences is laughably small. If you act rightly only if you do what in fact has the best consequences, then you can be fairly certain that you never (or almost never) act rightly.”¹³ And, or so it would appear, there is good

⁹Wiland, 355-356.

¹⁰Surely there are additional senses of “can” that entail that Jeffrey cannot cure cancer. See Howard-Snyder, 243-44. However, the sense of can I work with, and that seems most important for moral inquiry, is the sense of can that is implied by “did”: if I ϕ 'ed at t , it follows that I could have ϕ 'ed at t . See Miller, 53-54. Insofar as there is nothing stopping Jeffrey from having cured cancer at t —all he has to do is hit the right keys in the right order—it follows that he *can* cure cancer at t .

¹¹For a more pessimistic view of the good generated by a cure for cancer, see Albert Brooks’s futuristic novel *2030*.

¹²See Wiland, 356-7.

¹³Wiland, 359.

reason to reject consequentialism.

It is worth saying a little about what the objection in question amounts to. Wiland and Miller focus in on the suggestion that “[objective consequentialism] implies that our actions are *almost always* wrong.”¹⁴ But it seems to me that this cannot be the heart of the objection. There could be many reasons that our actions are almost always wrong: it could be, through some spontaneous feature of human psychology, almost all humans become moral jerks, and hence almost all humans almost always act wrongly. But we wouldn’t blame this on consequentialism. Rather, it seems to me that the objection runs somewhat deeper. The problem is that all humans almost always act wrongly even though they seem to be performing the best actions they can plausibly be morally required to perform. To put this another way, the fact of human cognitive limitations, intuitively speaking, seems to exempt certain actions from being an appropriate object of a moral requirement for a particular individual. The fact that Wilda doesn’t know how to open the safe seems to exempt her from a requirement to open it, even though she can, strictly speaking, open it by entering the right combination. But consequentialism seems to say that such actions *can* be morally required of Wilda and, by extension, Jeffrey. And so it seems to me that the objection, in essence, runs as follows: consequentialism requires actions that we think are inappropriate targets of a moral requirement, given the fact of human cognitive limitations, including procedural ignorance. The fact that virtually all human beings will act wrongly is merely a symptom of the disease of an incorrect theory of *moral requirability*.

2. *Human Cognitive Limitations and Non-Consequentialism*

As stated, the objection from human cognitive limitations is an objection to consequentialism. But the fact of cognitive limitations causes problems not simply for consequentialism but, as I shall argue, virtually every moral theory. Take the following principle:

Minimal Consequentialism (MC): that ϕ -ing will produce good consequences is a moral reason¹⁵ to ϕ , a reason that strengthens as the quality of the consequences of ϕ -ing increases.

Minimal Consequentialism is shared by consequentialists and many non-

¹⁴Wiland, 353; Miller, 55.

¹⁵Unless otherwise stated, I use “moral reason” to refer to objective moral reasons, reasons that determine the genuine moral valence of an action, rather than the moral valence of an action as assessed by a particular moral agent.

consequentialists. For consequentialists, MC is sufficient to fully catalog all moral reasons. However, a moral theory can accept *Minimal Consequentialism* without accepting consequentialism. After all, MC says only that good consequences of an action generate *a* moral reason for performance, not that this reason is overriding, or even comparatively very important. For instance, one could accept MC and a host of agent-centered restrictions, permissions, etc. However, any theory that accepts this general principle will face the problems noted in §1. Take Jeffrey’s case. Curing cancer by means of typing out the formula and widely disseminating it will produce overwhelmingly good consequences. But given that good consequences generate moral reasons, and given that the the good consequences of curing cancer are *overwhelming*, one would expect that no matter how comparatively unimportant the reason to promote good consequences is, the sheer magnitude of good caused by such an action would be enough to entail that the reason to do so is sufficient to morally require doing so. To put this another way, it would appear that (a) the enormity of the good involved should, on any plausible view, morally override any non-consequentialist permissions or restrictions and (b) even if not, it is very unlikely that typing out the cure for cancer would be particularly burdensome for Jeffrey (beyond slightly delaying the *denouement* of his gooseflesher), nor would it require him to commit an unpardonable moral sin. Even if the reason to promote goodness is very weak compared to all other moral reasons, a wide range of moral theories will face the problem of human cognitive limitations. Hence a wide range of moral theories must search for a solution.

One possible solution, whether for explicitly consequentialist theories, or for theories that treat consequences as at least one factor among many, is to reject the claim that moral reasons are a product of the actual consequences of particular actions.¹⁶ Rather, one might hold that moral reasons are generated by *expected* consequences, expected, presumably, by the agent in question. Correspondingly, such a view would accept:

Minimal Subjective Consequentialism (MSC): that ϕ -ing will produce good *expected* consequences is a moral reason to ϕ , a reason that strengthens as the quality of the expected consequences of ϕ -ing increases.

MSC suggests that the moral valence of my action is, at least in part, de-

¹⁶Wiland, for instance, explicitly focuses on what he calls “objective” consequentialism—a view that accepts MC rather than MSC, though he does not consider whether subjective consequentialism would be able to avoid the problem at hand.

terminated by the action's expected, rather than actual, consequences. MSC permits of varied interpretation; some will say that expected consequences are those that can be expected given the agent's actual epistemic states. Others will say that expected consequences are determined by what an agent would expect under certain idealized or counterfactual epistemic conditions.

Offhand, however, it is a little difficult to see why any view that accepts MSC rather than MC could avoid the objection from procedural ignorance.¹⁷ After all, the expected consequences of my curing cancer are certainly extremely good; the expected consequences of any other action I could perform at the time pale in comparison. If so, MSC is no fix: curing cancer not only has the best *actual* consequences, but also the best *expected* consequences. This is true even if we hold that expected consequences are determined from the perspective of the agent's actual epistemic states. As I sit, right now, I certainly have sufficient information to determine that the expected consequences of curing cancer are extraordinarily good. If so, or so it would seem, I have most reason to cure cancer, will not do so, and will fail to conform to my moral obligations almost all the time.

A response is worth considering. One might say that though the expected consequences of *curing cancer* are good, the expected consequences of taking the steps necessary to cure cancer, i.e., typing a set of random characters on one's keyboard, are not particularly good. After all, the chances that I will actually cure cancer by sitting down and typing out a bunch of random words or characters is pretty slim, to say the least. And if so, I can't be required to do so: the expected consequences are of very low comparative quality. But this response doesn't solve the problem. While I may not be required to sit down at my computer and type numbers randomly—which is a means to curing cancer—I am nevertheless required, given its foreseeable consequences, to *cure cancer*. To claim that I am not required to cure cancer given that I am not required to sit down at my computer and type a bunch of characters randomly is to treat as identical two distinct acts: the act of curing cancer and the act of typing a bunch of random keys.¹⁸ And insofar

¹⁷Dale Miller, in proposing a version of the problem discussed here, suggests that “even some utilitarian or non-utilitarian forms of probable-consequence act consequentialism” may succumb to it (Miller, 50).

¹⁸See, for instance, Alvin Goldman, *A Theory of Human Action* (Englewood Cliffs, NJ: Prentice Hall, 1970), chs. 1 and 2. I should note here that I will assume a fine-grained theory of action-individuation. This is, of course, controversial. But nothing substantive rides on this. One may also say that Jeffrey expects very good consequences to result from his action of “curing cancer by typing the right characters in the right order”. I will assume a fine-grained scheme in this paper, and will put arguments in terms of such a scheme, but the arguments can be translated to a coarse-grained scheme *mutatis mutandis*.

as we can distinguish between these actions, as we surely can, any view that accepts MSC will generate a moral reason to cure cancer. Thus MSC does not avoid the problems cited here.

And though the following point is tentative, it seems to me that even moral theories that refuse to treat expected or actual consequences as moral factors might have a very serious problem with the fact of human cognitive limitation of the sort I've been exploring. Imagine, for instance, that Jeffrey can cure cancer at very little cost to himself (which he surely can, despite not knowing how to do so). Under such conditions, it may very well be, e.g., a lack of respect for persons, or a failure to treat individuals as ends, or a lack of adequate beneficence, to fail to do so *insofar as one can*. But if this is correct, it would appear that *any* moral theory succumbs to the problem discussed here. As a general conjecture: whatever the factors specified by a given moral theory T, T will, on occasion, require actions that one can, but will not, perform simply given the fact of human cognitive limitations, and hence will offer far too wide a theory of moral requirability.

However, even if we allow that fully non-consequentialist moral theories can avoid these problems, MC and MSC are compatible with a very wide range of moral theories. Any moral theory that cares about the consequences of our actions will accept one or the other. Insofar as there appears to be very good moral reason to care about consequences (even if just a little bit), a fix for the problem of moral obligation in light of human cognitive limitation is thus of the first importance not just for consequentialism, but for moral theory more generally.

3. Moral Requirability

One cannot avoid the problems cited here simply by rejecting consequentialism. One could avoid it by rejecting both MC and MSC (assuming my conjecture is false). But this proposal is sufficiently unattractive as to warrant exploring other alternatives.

Fortunately, MC and MSC are not by themselves the source of the problem. To see this, allow me to introduce a bit of terminology. Call an action that is eligible to be the object of a moral obligation for a person x at time t a "requirable" action for x at t . The set of morally requirable actions is a substantive matter, but need not differ among moral theories. A strict Kantianism and a strict act-utilitarianism will certainly differ concerning what morally requirable action, for x at t , is morally required. But they needn't differ about which actions are *requirable*. A second point is worth mentioning. Theories of moral requirability imply not only limitations on

those actions that are morally requirable, but also limits on the the reach of *moral reasons*. This follows from the relationship between moral reasons and moral requirements. If R is the set of morally requirable actions for x , and there are moral reasons for actions that are not members of R , it could be that in some cases, x has strong, even optimal, moral reason to perform an act that is not a member of R . But if so, it would seem plausible to declare that one is morally required to perform that action, voiding the theory of moral requirability on offer.¹⁹ Hence it must be that the set of morally requirable actions and the set of actions for which there can be moral reasons is identical.

My claim is that moral theories succumb to the problem of human cognitive limitations not simply given their embrace of MC or MSC, but their embrace of MC or MSC (*modulo* my conjecture) along with the following theory of moral requirability:

The Traditional View (TRAD): ϕ is requirable for x at t if and only if x can ϕ at t .

TRAD, as a theory of moral requirability could be—and plausibly is—shared by virtually all substantive moral theories. TRAD does not say how one ought to evaluate the actions it identifies as morally requirable. Rather, it simply restricts the set of actions one could be under a moral requirement to perform. Pursuant to the above paragraph, TRAD also restricts one’s moral reasons: there are no moral actions one cannot perform.

Just as there are good reasons to accept or reject substantive theories of moral requirements, there are good reasons to accept or reject individual theories of moral requirability, no matter what substantive account of the moral valence of requirable actions one accepts. For instance, imagine the following view:

The Hard View (HARD): ϕ is requirable for x at t if and only if there is a possible world in which x ϕ s at t .

The Hard View, unlike TRAD, claims that one can be morally required to perform actions one cannot perform: all actions that are *metaphysically possible* for x to perform at t are morally requirable for x at t . However, HARD suffers in comparison to TRAD. And the major problem has already been explored. If actions that are merely metaphysically possible for me

¹⁹I have assumed for the sake of argument here that moral rationality is optimizing. But the same problem arises even for a satisficing view: it could be that the only action for which there is *enough* moral reason is not a member of R .

can be morally required of me, moral behavior is virtually impossible. After all, HARD allows that I can be required to perform actions that I cannot perform in any possible world that contains the same set of natural laws as the actual world. Surely, however, were I to leap tall buildings in a single bound, or cure cancer at the snap of my fingers, or cast spells of happiness for all and sundry, doing so would produce better consequences than any action I *can* actually perform. And so, if we accept HARD, doing so is morally required. But this is an implausible result. Surely, as we have already seen, I act in a morally justifiable manner more often than this.

We are not conceptually bound to accept TRAD. Indeed, some wider theories of moral requirability have been floated in the moral and political philosophical literature. Consider, for instance, the following suggestion from Martha Nussbaum. Nussbaum argues that it is a moral obligation of all societies to provide all individuals a threshold level of ten “basic capabilities.” She also says, however, that: “In desperate circumstances, it may not be possible for a nation to secure them all up to the threshold level, but then it becomes a purely practical question what to do next, not a question of justice. The question of justice is already answered: justice has not been fully done here.”²⁰ Here Nussbaum seems to claim that even if particular societies cannot actually bring all individuals up to the appropriate level they have failed to conform to their moral obligations; the obligation Nussbaum cites does not wait upon “can”. I think, and have argued elsewhere, that Nussbaum’s claim here is implausible.²¹ But it is not implausible because she is *conceptually* mistaken; she does not misuse the terms “justice” or “moral obligation”.²² Rather, her errors are in adequately accounting for our considered judgments concerning the limits of moral obligation. Nussbaum allows that x (in this case a political society) can be morally obligated to perform acts that it is not plausible to treat as requirable for x .

4. Alternatives

If TRAD is not a feature of the concept of moral obligation, alternatives are

²⁰Martha Nussbaum, *Frontiers of Justice* (Cambridge, MA: Harvard University Press, 2006), 175.

²¹Dale Dorsey, *The Basic Minimum: A Welfarist Approach* (Cambridge: Cambridge University Press, 2012), ch. 4.

²²One might dispute this, of course. However, even if we accept that TRAD is conceptually correct, there still remains a question concerning how we interpret the “can” in TRAD. One might then interpret my alternative proposal simply as a particular theory of the way to understand this crucial modal verb that adequately captures its moral significance.

worth exploring. Because TRAD allows that one can be morally required to perform any action one can perform, substantive theories of morality will assess every action one can perform for its comparative moral valence. But this yields that, for Jeffrey, curing cancer will be morally required: he can cure cancer, and on virtually any theory that accepts MC or MSC doing so will have a favorable moral valence in comparison to any other action he can perform. Hence it would seem sensible, if we are concerned about the puzzle of human cognitive limitations, to narrow the range of morally requirable actions that TRAD accepts, just as TRAD narrows the range of morally requirable actions that HARD accepts.

But how? Examining the cases of Wilda and Jeffrey might lead to an obvious possibility:

The Knowledge View (KNOW): ϕ is morally requirable for x at t if and only if x knows how to ϕ at t .

Insofar as Wilda does not know how to open the safe, and Jeffrey does not know how to cure cancer, it would appear that KNOW entails that they cannot be morally required to do so even given the quality of the resulting consequences. So far so good.

However, there are good reasons to reject KNOW. KNOW it is both under-inclusive and over-inclusive.

First, it is under-inclusive. Consider:

Sondra: Sondra stands before a safe that, were she to open it, cancer would be cured. Also imagine that Sondra believes that entering the numbers 867-5309 will open the safe. Furthermore, her belief is justified and true—this number is, in fact, the safe’s combination. However, imagine that Sondra’s justified true belief was a product of a Gettier-like scenario,²³ and hence that this particular number will open the safe does not count, for Sondra, as knowledge.

In this case, Sondra does not *know* how to open the safe. Her justified true belief is not knowledge. Nevertheless, it would seem absurd to say that she could not be required to open the safe. After all, she believes she knows how to do so, that belief is justified, and, were she to actually enter the number

²³Imagine, for instance, that Sondra overheard the owner of the safe declare that anytime he had a safe, he would set the combination as the title of his favorite Tommy Tutone song. But also imagine that the owner never actually set the combination, and it just happened to have been set to that number randomly in the factory.

she believes will open the safe, cancer would be cured. Not only can she be required to open the safe, her failure to do so appears, to me anyway, to be a grave moral wrong.

Second, KNOW is over-inclusive. Some actions we know how to perform are inappropriate targets of moral obligation given *other* cognitive limitations we might possess. Assume, for instance, that the correct combination to Wilda's safe is 867-5309. Wilda certainly *knows how to enter* 867-5309. Insofar as she knows how to work a safe, she knows how to enter any particular number into that safe.²⁴ But requiring Wilda to enter 867-5309 is certainly no more plausible than requiring Wilda to *open the safe*.²⁵ Parallel reasoning applies in Jeffrey's case. We can assume that there is some string of characters, *S*, such that were Jeffrey to type *S*, he would cure cancer. Insofar as Jeffrey knows how to type, he knows how to type *S*. KNOW implies that Wilda can be morally required to enter 867-5309, and Jeffrey can be morally required to type *S*. But it seems implausible to say that, e.g., Wilda should be required to do so; the reason surely is, at least in part, that she fails to recognize entering 867-5309 as a method by which to open the safe.²⁶

A further possibility should be considered here. Take:

The Counterfactual Intention View (INTEND): ϕ is morally requirable for x at t if and only if, were x to intend to ϕ at t , x would ϕ at t .

This view seems to avoid holding Wilda and Jeffrey morally accountable for failing to open the safe, and failing to cure cancer, respectively. Given their cognitive limitations, we can assume that, were they to intend to cure cancer or open the safe, they would fail to do so. INTEND also avoids the problematic suggestion that Sondra cannot be morally required to open the safe. However, INTEND also seems to imply that Wilda can be morally required to enter 867-5309, and Jeffrey can be morally required to type *S*.

²⁴Cf. Carlson, 92.

²⁵Note, again, that I am treating these actions as distinct. However, if entering 867-5309 and opening the safe are the same actions (under different descriptions), it is clear that Wilda knows how to open the safe, because she knows how to enter 867-5309. And hence KNOW fails to solve the original problem. I will assume, however, that these are different actions—partly on grounds that it seems wrong to me to say that simply because Wilda knows how to enter 867-5309 she thereby knows how to open the safe—for the sake of convenience.

²⁶Admittedly, this might not be a problem for KNOW on the assumption of MSC. It is certainly the case that entering 867-5309 does not have very good expected consequences. Nevertheless, KNOW seems to me to fail for two reasons, and the former is applicable even if we accept MSC rather than MC.

Were they to intend to perform those actions, they would do so. However, an important response arises for INTEND that is not available for KNOW: if we conjoin INTEND to MSC rather than MC, we can hold that, because the expected consequences of entering 867-5309 are not particularly good, INTEND+MSC avoids the problematic claims we've been discussing so far.

However, I remain unconvinced that INTEND can solve the problem at issue in a satisfactory way. Consider, again, Jeffrey. Imagine that it *just so happens* that *were* Jeffrey to intend to cure cancer, he would enter the right characters into his computer, and would, in fact, succeed at so doing, despite the fact that whether he does so or not is simply a one-in-a-trillion shot. Would we say that Jeffrey behaved in an immoral manner in failing to do so? I think this result should be especially troubling for partisans of MSC. Here Jeffrey appears to be morally required to ϕ , whereas some other person, Jackie, is not morally required to ϕ , simply given the fact of an ability on Jeffrey's part about which he has no evidence, and for which there could be no evidence in any sensible idealized counterfactual set of circumstances. Hence it seems to me that INTEND+MSC yields results that are in tension with the general motivations of a subjectivized approach to consequentialist moral reasons. Of course, this verdict is not in tension with MC. But INTEND+MC is over-inclusive in just the way KNOW is.

These examples show that the particular human cognitive limitation involved is not itself procedural ignorance—it is not ignorance of how to perform a particular action.²⁷ Rather, the cognitive limitation is something else. But then what is it? Take again the action of entering 867-5309 into the safe. Why should we think that Wilda cannot be morally required to enter that number? It seems to me that the reason we would refrain from treating this action as morally requirable for Wilda is not an epistemic limitation, but rather a *deliberative* limitation. Wilda cannot see this particular action as having the sort of properties that would give rise to a reason to perform it.

To put this in a slightly different way, Wilda's own capacity as a deliberative agent is insufficient to bring it about that she would, in fact, enter the correct number into the safe. And while I will say more about my preferred alternative in the next section, I introduce it here as:

²⁷Furthermore, it seems to me that this example shows something else: that the problem of human cognitive limitations cannot be solved by accepting that TRAD is analytic, and simply mucking about with the proper sense of 'can'. (See note 19.) In any sense of 'can', Wilda can enter 867-5309. But it is unacceptable to require her to do so. And so TRAD is surely not analytic, and the problem of human cognitive limitations cannot simply be solved by finding the proper sense of 'can'.

The Agency View (AGENT): ϕ is morally requireable for x at t if and only if x can ϕ at t as an agent.

4.1. *AGENT and Human Cognitive Limitations*²⁸

To see my proposal in more detail, and why it offers a solution to the problems I investigate here, I must first explore the notion of “agency” with which I seek to work. My intention is not to provide a full theory of agency or any cognate concepts. I remain neutral on such topics as whether moral deliberation is a result of cognitive, conative, or affective pro-attitudes, the nature of autonomy or autonomous action, whether deliberative agents act under the guise of the good, or whether normative judgments are necessarily motivating, etc. Rather, I seek a minimal theory of what it means to act as an agent, one that I hope will be broadly ecumenical when it comes to substantive controversies such as those noted here.

In assessing whether a particular action can be performed as an agent in the sense I mean here, it is insufficient to say that the action was (a) performed by x and (b) x is a moral agent. Moral agents can sometimes fail to perform actions *as* agents, at least in the sense I mean. For instance, moral agents simply decide based on whim or caprice, they might govern their actions by the simple flip of a coin. None of these decisions entail that they are not moral agents. But they do entail that the resulting actions are not performed *as agents*. Rather, AGENT holds that morally requireable actions are not those performed *by* agents, but those that are performed *as a result of* normative deliberation or agency. To perform an action as an agent is to perform that action on the basis of one’s own deliberative agency. It is, in other words, to *see a reason to perform that action*, and to *perform the action on the basis of that reason*.²⁹

Two qualifications should be made clear immediately. To perform a particular action as an agent, it is insufficient simply to see *some reason or other* to perform that action. The reason has to be of the right kind. To see what I mean, take the following case. Imagine that I stand at a roulette table, and I see reasons to play roulette: it will be thrilling, it will satisfy a desire of mine, etc. However, I see no reason to bet on any *particular* number. So I bet on 22.

²⁸This section was improved by helpful conversations with Julia Driver.

²⁹Again, this is neutral between substantive theories of normative deliberation or agency. Reasons involved may be given by one’s rational capacities, or they may simply be desires; furthermore, “normative deliberation” could involve simple inquiry into one’s own desires, or it could involve some essential rational capacity. I’m neutral on all of these topics for present purposes.

In this case I have performed a number of different actions. I have played roulette. I have bet on 22. I have bet on a black number, an even number, and a number in the second third of the table. However, if we ask: “have I bet on 22 as an agent?” it seems to me the answer is clearly *no*. Of course, I have played roulette as an agent, insofar as I deliberated on the basis of reasons I recognize to play roulette.³⁰ But I haven’t bet on 22 on the basis of my deliberative agency. This is true despite the fact that I certainly see some reasons to bet on 22. Insofar as betting on 22 is a way to play roulette, this is certainly a reason I will recognize to bet on 22. Insofar as betting on 22 is one way to not cut my own legs off, which I’d prefer to avoid doing, this is certainly a reason I recognize to bet on 22. But importantly none of these considerations that I see will tell in favor of betting on 22 rather than placing any *other* bet on the roulette table. The reasons I recognize support not betting on 22, but rather a disjunction of potential actions I might perform.

Thus to perform an action as an agent it is not sufficient to see some reason or other to perform the action. Rather, one must see what I shall dub a “contrastive” reason. A contrastive reason to ϕ is a reason to ϕ *rather than* any other action ψ one might perform. The reason(s) one sees cannot simply support a disjunction of potential actions. They must support *that action in particular*.³¹

The second qualification runs as follows. To ϕ as an agent, it is not sufficient to ϕ on the basis of some contrastive reason I see to ϕ . Rather, it must be the case that the difference between ϕ -ing and merely *trying* to ϕ is under one’s deliberative control. Take the following example:

Sarah: Sarah, a scientist, believes that there are good reasons to produce organism X in a certain petri dish. However, though this person knows that she must, at the very least, leave open

³⁰Again, this presumes a fine-grained theory of action, viz., that the action of playing roulette and the action of betting on 22 are distinct. But if the coarse-grained theory of action is correct, one might rephrase AGENT by holding that one is required to perform an action under a certain description if that action, under that description, can be performed as an agent.

³¹One might say that this radically restricts the range of actions that I might perform as an agent. Imagine, for instance, that I recognize a reason that tells in favor only of *two* potential actions ϕ and ψ ; though it rules out all others I might perform. On this view, if I see no reason to ϕ rather than ψ I cannot ϕ as an agent, despite the fact that I might deliberate, and on the basis of my deliberation, decide to either ϕ or ψ . But it seems to me that this verdict is exactly the correct one. If I cannot see a reason to ϕ rather than ψ , we might say I “decided, as an agent, to ϕ or ψ ”. If we allow the possibility of disjunctive acts, we may say I ϕ -or- ψ ’ed as an agent. But we would not say that I ϕ ’ed as an agent.

the petri dish, she doesn't know what to add to it, where to place the petri dish, at what temperature the petri dish must be kept, or any other details about the process by which organism X is grown. On a whim, she places it at a certain spot and temperature, adds random chemicals, and leaves for the night. When she returns in the morning, she has grown organism X.

We would say that Sarah grew organism X, surely. But we would not say that she grew organism X *as an agent*, as a result of her deliberative agency? Surely not. This is true despite the fact that she sees contrastive reason to grow organism X, and, one can assume, threw a bunch of chemicals together partly on the basis of this contrastive reason. How, then, to explain what Sarah lacks?

To explore this point in more detail, it is helpful to make use of the idea of a “longish” action.³² Some actions unfold over time, and are made up of individual sub-acts. Sarah has deliberative control over some of the individual sub-acts and not others. Hence we can make a distinction between the sub-acts of growing organism X that Sarah can perform as an agent, and those sub-acts that Sarah cannot perform as an agent. Some of these sub-acts, however, spell the difference between trying to grow organism X and actually succeeding at doing so (which she does, but not as an agent). Sarah has deliberative control over whether she washes up, uses a petri dish rather than a dirty lunchbox, places it in a temperature controlled incubator rather than in a hot iron skillet, etc. These are all sub-acts for which Sarah sees contrastive reason. Nevertheless, these sub-acts are not sufficient to succeed at growing organism X. The petri dish must be placed at the *right* temperature, with the *right* chemicals. But Sarah can see no reason to select the right chemicals rather than any other set of (wrong) chemicals.

Thus my account of what it means to perform an action as an agent runs as follows. One ϕ s as an agent to the extent that (a) one ϕ s, (b) one sees contrastive reason to ϕ , and (c) one sees contrastive reason to perform any individual sub-acts that constitute the difference between merely *trying* to ϕ and ϕ -ing. Once these conditions are in place, one ϕ s as an agent.

One note before I move on. Some will complain that my account of what it means to ϕ as an agent does not match their considered judgments. Fair enough. However, I argue in the next section that there is independent reason to believe that morality should not treat as requirable actions that could not be performed as an agent *in my sense*. If that's right, then not much will ride on whether my account of the performance of an action on

³²Goldman, ch. 2.

the basis of deliberative agency is the true account. Readers can feel free to treat my account as stipulative, as an account of “schmagency”, as it were, if they so choose. The reasons in favor of AGENT (understood in my way) will remain the same.

That AGENT avoids the problem of human cognitive limitations should be relatively clear. The basic idea runs as follows. Recall that it seems implausible to say, of Wilda, that she can be morally required to enter 867-5309 and to open the safe. And so to solve the problem of human cognitive limitations, AGENT must be able to say of Wilda that entering 867-5309 and opening the safe are not requirable of her. Take, first, entering 867-5309. In this case it is relatively clear that Wilda cannot see any contrastive reason to do so. Were she actually to enter the right number, she would not do so on the basis of normative deliberation, but simply as the result of a guess, a random selection, etc. Parallel reasoning applies in the case of Jeffrey’s moral requirement to type *S*.

Take now Jeffrey’s purported moral requirement to cure cancer. It might seem that AGENT cannot avoid morally requiring Jeffrey to cure cancer: though he sees no reason to type *S*, he *does* see reason (even contrastive reason) to *cure cancer*. But if that’s right, then it would seem that Jeffrey, though he cannot be morally required to type a certain set of characters, he can, and is, morally required to cure cancer.

However, it seems to me that though Jeffrey can see reasons to cure cancer, he cannot do so as an agent. Recall that, as in Sarah’s case, to perform the action as an agent, one must see a contrastive reason to perform the individual sub-acts that spell the difference between trying to ϕ and ϕ -ing. In the case of curing cancer, the action unfolds over a period of time. But insofar as Jeffrey is unaware of the proper procedure to cure cancer, he is unable to see reasons to take the *proper* procedure rather than some alternative *improper* procedure. When Wilda is faced with a safe the combination to which she does not know, she cannot unlock the safe as an agent, because she can see no reason to enter 867-5309 rather than 438-2368 rather than any other combination of numbers. This point generalizes. Because Jeffrey hasn’t the foggiest idea how to cure cancer, it seems hard for him to properly *cure* cancer as an agent. If, against all odds, he actually cures cancer he will not have done so as an agent. He will have *tried*, perhaps, to cure cancer as an agent, but—as in Sarah’s case—he will not have seen contrastive reason to perform the crucial sub-acts that constitute curing cancer. His performance of them would have been the result of mere guesses. Parallel reasoning applies in the case of Wilda’s moral requirement to open the safe.

Of course, Jeffrey could deliberate, decide that the action of curing cancer is supported by weighty reasons, and therefore, for this reason, *punch random keys*. If we accept AGENT and MC (but not, plausibly, AGENT and MSC), this will be morally required. But whether or not this is a plausible result just seems to me part of the more general dispute between those who accept MC and those who accept MSC. Whether one accepts MC or MSC, however, AGENT provides more plausible results than alternative theories of moral requirability. Given that morally requirable actions are fixed to those that an individual can perform as an agent, i.e., as a result of normative deliberative agency, human cognitive limitations are not a factor in determining whether an individual will or will not perform an action that can be, under AGENT, morally required. After all, each action that, for a particular individual, can be morally required can be brought about by the operation of a moral agent's deliberative capacity. Whether this can be done by any one person is surely determined in part by that person's cognitive limitations.

5. Objections

In the final section of the paper, I consider five important objections to AGENT. The first argues that AGENT is *ad hoc*, especially in light of the specialized theory of what it means to act as an agent that I offer in the previous section. The second holds that AGENT is in tension with consequentialism. The third argues that my view unduly restricts the range of objective moral reasons. The final two argue (though in different ways) that AGENT is under-inclusive: it rules out the possibility of morally requiring particular actions that really should be requirable.

5.1. Ad Hoc

Perhaps the most important objection to AGENT runs like this: why believe that this account of agency has anything to say about what is or is not morally requirable? Even if, or so it may be claimed, AGENT is successful at turning back the problem of human cognitive limitations, what independent reason is there to believe it? Isn't this simply an *ad hoc* addition to our understanding of the boundaries of moral requirement?

There are two points I offer in favor of AGENT here. The first can be illustrated by means of the following example:

Susan: Susan is walking home from the bookshop one day, when

she happens to catch a glimpse of a child in the way of an oncoming truck. If she acts quickly, she can save the child. If she delays for even a split second, however, she will be unable to save the child.

If TRAD is correct, we can certainly assume that, for Susan, saving the child is morally required. After all, she can do it. Furthermore, this child is no would-be Hitler, saving the child would cause no other moral problems, etc. But should we say that Susan is morally required to save the child? Imagine, for instance, that Susan sees the child and hesitates for a split-second, at which point it is too late. Would we say, of Susan, that she behaved immorally? Certainly not! Imagine that Susan instead sees the child, and acts, solely on instinct, to save her. Would we say that she conformed to her moral obligations? Again, I find this suggestion odd, even perverse. In describing her action, I find it much more natural to use axiological or evaluative terms: “it’s good/fortunate/a miracle that she acted”, rather than “she acted as she ought to have acted/she fulfilled her duty/she acted rightly”. I think our (or at least my) reaction to this case can be traced to the extent to which Susan could have saved the child or not on the basis of normative deliberation. The fact that there was—quite literally—no time to think seems to me to move saving the child beyond the reach of moral requirement. If it’s the case that a particular action ϕ could be performed by a particular person x only as a result of luck, instinct, or fortune, we should refrain from describing this sort of action as morally required.

But it is important to see that for a particular person to be able to ϕ in a way that does not depend on luck, instinct, or fortune, it is required that this person see *contrastive* reason to ϕ . If a person can see no reason to ϕ rather than to ψ , it would seem that whether he or she will ϕ rather than ψ will be a product of happenstance in just the way that Susan’s grabbing for the child is a product of happenstance. Parallel reasoning applies to the requirement that to perform an action as an agent, individuals have deliberative control over the distinction between trying to ϕ and ϕ -ing successfully. If one does not have such control, whether one actually ϕ s rather than tries to ϕ will be a product of the sort of luck, instinct, or fortune that seems to defeat a moral requirement to ϕ .

Next, consider the following case:

Joey: Joey is the victim of a dastardly neuroscientist, who has implanted a device in his brain. This device shuts off Joey’s ability to act as an agent during one half-hour window on Thursday

evenings. Before and after this period, his deliberative capacity operates in perfect working order.

Imagine that Joey kicks someone in the shins on Thursday evening. Would we be tempted to say that Joey behaved in a way he morally ought not to have? If Joey has the ability to, say, perform some good deed on Thursday evening, but does not, would we say that Joey has failed in his moral obligations? I find this suggestion absurd. However, it seems to me that things change after that half-hour window ends. If, at 5:45 on a Tuesday, Joey can decide whether to rescue a drowning child or to protect his new suede shoes, it seems to me we would say that Joey is morally required to save the child (or save his shoes, depending on the moral factors involved). In this set of circumstances, Joey has deliberative control over his actions; whether he behaves in one way rather than another is not determined simply by luck, happenstance, or caprice. But if this is correct, AGENT is further confirmed: the distinction between Joey-on-Thursday and Joey-at-any-other-time is the distinction between an individual who cannot, and who can, perform actions as an agent. And so it would appear that the bounds of moral requirability track our ability to deliberate and to perform actions on the basis of normative deliberation. AGENT is not *ad hoc*.

5.2. *Tension with Consequentialism*

Theories of moral requirability should be ecumenical between substantive theories of the moral valence of requirable actions—or, at least, as ecumenical as possible. Hence a sensible objection to any theory of moral requirability is that it is in tension with substantive theories of morality that may themselves be plausible methods by which to assign moral valence. But, or so it may be claimed, AGENT is in tension with consequentialism. To see this, consider a number of consequentialist platitudes. For consequentialism, the consequences of ϕ determine ϕ 's moral valence *period*. One needn't deliberate in any particular way, one needn't even deliberate at all. Rather, consequentialism simply declares that one's actions are morally acceptable if and only if they produce the best possible consequences. Consequentialism has a built-in distance from the deliberative capacity of agents; if so, the general structure of consequentialism is in tension with any limiting condition that makes explicit reference to deliberative agency.

But this objection misunderstands AGENT *qua* theory of moral requirability. AGENT does not say that the *content* of deliberation or the particular reasons chosen, or indeed whether deliberation was engaged in at

all, makes any difference to the moral valence of x 's performance of a given act ϕ . Indeed, AGENT is entirely neutral with respect to the moral valence of requirable actions. Rather, it offers only a limiting condition: instead of insisting that x 's moral requirements be chosen from those actions x can perform, it says x 's moral requirements must be chosen from those actions that x can perform *as an agent*. The moral valence of actions within this set is not assessed by AGENT, and hence AGENT says nothing that is or could possibly be in tension with consequentialism or any other substantive theory of morality. AGENT simply says that *however* moral theories are to evaluate actions, the actions they evaluate must be open to a given individual as an agent. AGENT is thus not in tension with consequentialism: the set of actions that could potentially be morally required could certainly be rank-ordered in a consequentialist way.

However, there may be a further, perhaps more obvious, problem at work here. Some might hold that TRAD is *cooked into* the consequentialist theory of right action. Consequentialism, on this view, would simply hold that one is morally required to perform the action, among those one can perform, that has the best consequences. Of course, AGENT is not compatible with this view. Good for AGENT: this way of understanding consequentialism renders consequentialism defeated by the fact of human cognitive limitations. However, AGENT can retain the central consequentialist claim that the moral quality of actions are a function, solely, of the axiological quality of their consequences. This position seems to me to capture all of that which renders consequentialism attractive without the troubling result that decent folks never act rightly.

5.3. *Objective Moral Reasons*³³

Third, one might think that AGENT unduly limits the range of moral reasons that apply to a given person. Take, for instance, the following case:

Roger: Roger is a member of a crack bomb squad and is in the middle of defusing a bomb that, if left undefused, will destroy a large office building, killing dozens in the process (as well as himself). Roger knows that to defuse the bomb he must either cut the red wire or the green wire. But he has no idea which wire he should cut. He can see no reason to cut the red wire rather than the green wire, or *vice versa*. As it happens, the red wire defuses the bomb.

³³Thanks to Doug Portmore for Roger's case, and a very helpful exchange on this topic.

Several things seem natural to say about this case. First, given that the red wire defuses the bomb, it may seem plausible to say that Roger morally ought to cut the red wire. In addition, it seems natural to say that, despite his lack of evidence, there is an *objective* moral reason for Roger to cut the red wire: doing so will save dozens of lives.

The problem appears to be that on AGENT none of these things come out true. Insofar as Roger cannot see a reason to cut the red wire rather than the green wire, it would appear that Roger can do neither as an agent: his choice of red or green will not be on the basis of a reason in favor of one rather than the other. He could, of course, decide to cut *one of the wires at random* as an agent (just as a person could, as an agent, decide to type random keys on a keyboard). And, as a result of his random choice, he may end up cutting the red wire. But he sees no reason to cut the red wire rather than the green wire, and hence cannot be morally required to cut the red wire. But, first, to say that Roger has moral reason to cut the red wire seems rather commonsensical. Second, in being unable to say that Roger has no moral reason to cut the red wire, it would appear that AGENT renders objective moral reasons indistinguishable from subjective moral reasons. After all, one important feature of objective moral reasons is that they apply to particular individuals no matter what their *assessment* of the moral reasons that apply to them happen to be.

I'll take these in reverse order. It is surely not correct that AGENT fails to allow a distinction between objective and subjective moral reasons. AGENT does, of course, restrict the range of actions a person can be said to have objective moral reason to perform, and this range is determined in part by a certain sort of fact internal to the agent in question, i.e., whether this agent can see a reason to perform that action. But this does not erase the distinction between objective and subjective moral reasons. For starters, depending on one's preferred theory of subjective reasons, to *see* a moral reason to ϕ does not guarantee that there is, in fact, subjective moral reason to ϕ . It could be, for instance, that one's subjective moral reasons are culled not from that which one believes one has objective moral reason to perform, but on the basis of what one would believe one has subjective moral reason to perform if one were in, say, some favored or idealized epistemic state. Leaving this aside, however, even if to see a moral reason to ϕ is to have a subjective moral reason to ϕ , this does not erase the distinction. First, it could be that one sees a reason to ϕ , but only a prudential reason to ϕ . If so, one can have objective moral reason to ϕ even if there is no *subjective* moral reason to ϕ . Furthermore, the subjective moral significance of ϕ -ing—which is certainly determined in part by an individual's actual or counterfactual

mental states—could be substantially different than the objective moral significance of ϕ -ing. Hence there remains a robust distinction, according to AGENT, between objective and subjective moral reasons.

However, the real issue here is not whether AGENT can continue to make a meaningful distinction between objective and subjective moral reasons, but rather whether its verdicts in Roger’s case are wildly unintuitive. Certainly we would like to say that Roger has objective moral reason to cut the red wire. I humbly admit that this seems, at first glance, like the right answer. But should we treat this as dispositive reason to accept TRAD rather than AGENT? Consider the following. Imagine that Roger faced not only two wires, but 1000. It seems to me that there is no principled reason to say that Roger has objective reason to cut the red wire in the case above, but doesn’t have objective reason to cut the 456th wire (which is the correct one) in the latter case. But if this is correct, Roger’s case simply becomes a version of Wilda’s, and hence to suggest that Roger has an objective moral reason to cut the red wire seems committed to the claim that Wilda has an objective moral reason (and hence objective moral requirement) to enter 867-5309. Hence it seems to me that to accept Roger’s objective reason to cut the red wire is to commit to a theory of moral requirability that we should reject for reasons already explored. This is, of course, not to gainsay the contrary intuition. But I think we should regard this considered judgment as a remnant of a theory of moral requirability that, on reflection, we do better to reject.

6.3. *Optimal Non-Agency*

The remaining two objections to AGENT pursue the joint suggestion that AGENT is under-inclusive. The first such objection runs like this. Consider the set of all actions (A) x can perform at a given time t . The set of actions that this same individual can perform at time t *as an agent* is a proper subset (R) of A . Now imagine that, at t , ϕ , a member of R , is morally required, insofar as ϕ is the morally optimal member of R . Imagine now that x fails to ϕ . It would appear, in failing to ϕ , that one fails to perform an action that is morally required, and hence behaves immorally. However, imagine that instead of ϕ -ing, x performs a further action, ψ , a member of A , but not a member of R . Furthermore, stipulate that, of the actions in A , ψ -ing is optimal with respect to the quality of its consequences (and, furthermore, there are no non-consequentialist moral reasons to prefer ϕ to ψ). It would appear that under these circumstances, one acted immorally in performing an action that, in fact, is optimal in comparison to ϕ . But this is absurd! Not

to put to fine a point on it, but we would certainly not declare that Susan behaved in a morally impermissible manner if she—without thinking—acted to save the child in question rather than continuing to walk home from the bookshop, which would, for her, be the optimal act she could perform as an agent.

This is an important and penetrating objection. I offer two mutually exclusive responses for those who are unwilling to simply bite the bullet. First, those who hold that it is implausible to describe Susan’s saving of the child as morally impermissible might appeal to an important distinction in kinds of moral reasons, viz., that between moral *requiring* reasons (i.e., reasons that count in favor of a moral requirement to ϕ), and moral *justifying* reasons (i.e., reasons that count in favor of a moral justification to ϕ). Moral justifying reasons have no power to require any actions; they simply have the power to justify certain actions that would otherwise be unjustified.³⁴ Insofar as AGENT restricts moral requirability, it need not restrict moral justification, and hence we might say that a theory of moral requirability limits only the range of moral requiring reasons, not moral justifying reasons. Hence we might allow that *justifying* reasons can apply to all members of A , and that depending on the strength of these justifying reasons, acts that do not fall within R can be morally justified. Of course, it remains the case that in being optimal with respect to the other members of R , ϕ is morally required *rather than* any other members of R . But this does not entail that it is required *tout court*. In particular, there could be other morally justified actions, including ψ , which fall outside R . But this does not change the central thesis of AGENT, viz., that only actions that fall within R can be morally requirable. Which actions that do not fall within R are morally justified is, of course, a substantive matter, and will be determined in large measure by the first-order moral theory one accepts.³⁵

There is, however, a second response, one I prefer. This proposal holds, in essence, that actions falling outside the bounds of moral requirability maintain no moral status at all. If we accept this claim, we can escape commitment to the implausible view that Susan’s rescue of the child in question is morally impermissible. To see this, note that the objection on the table holds that AGENT implies that, because ϕ -ing is morally required,

³⁴See Douglas Portmore, *Commonsense Consequentialism* (Oxford: Oxford University Press, 2011), ??.

³⁵For consequentialists, it may be sensible to declare any action with as good or better consequences than the optimal member of R is morally justified; in this way, one never acts in an impermissible way if one performs an action the consequences of which are at least as good as the optimal member of the morally requirable set.

and ψ -ing is non-identical to ϕ -ing, ψ -ing is immoral. However, there are two ways of failing to conform to a morally required action. One could perform an *immoral* action. Alternatively, one could perform an *amoral* action, an action that is beyond the bounds of moral assessment. If we admit that some actions are beyond the bounds of moral assessment, we can say that it is not the case that failing to conform to morally required action is always morally impermissible.

One might think that this proposal is incoherent. After all, if a particular action is morally *required*, that means morality entails that one *must* perform this action. And if one does not perform this action, surely one behaves in a morally unacceptable way. Though I discuss this puzzle in more detail elsewhere, I offer one (very sketchy) way around here.³⁶ One can allow that actions that do not fall within R lack a moral valence if we understand moral obligations as relative to a contrast class, i.e., relative to other actions that might potentially have a moral valence. If one is morally required to ϕ , we should understand this as being morally required to ϕ *rather than* any other action that falls within R . Hence to perform an action that is not morally evaluable is not to fail to conform to one's moral requirements: one is morally required only to ϕ *rather than* the other members of R . This proposal allows us to say, which seems plausible, that one behaves in a morally unjustified way only insofar as one performs an action that (a) is non-identical to a morally required action *and* (b) performs an action that is susceptible to moral assessment. As we have already seen, actions like Susan's seem inappropriate targets of moral assessment. And if this is correct, though Susan was morally required to walk home from the bookstore, and did not do so, this does not entail that she acted in a morally inappropriate way in saving the child. Rather, because moral assessment ceases at the bounds of R , she performed an action that has no moral valence whatsoever.

6.4. Bob

Finally, AGENT might be under-inclusive for the following reason. Consider:

Bob: Bob is a member of a hard-core Satanic cult. Bob believes that allowing drowning children to die is morally required, and seeks to subject children to drowning whenever possible. In fact, Bob's Satanic indoctrination is such that Bob can see no reason not to drown children.

³⁶Dale Dorsey, "Amorality: A Puzzle and Solution", MS.

Now imagine that Bob is faced with the choice of costlessly saving a child from drowning, or letting that child drown. Most would say that, despite the various attitudes relevant to Bob's normative deliberation, Bob is morally required to save the child. But the question now arises: given the structure of Bob's attitudes, could he save the child *as an agent*? Not on my account. After all, *given* his attitudes, he will not see contrastive reason to save drowning children. If he ends up saving a drowning child, he will not be doing this as a result of deliberation—given his psychology—but simply as a matter of mere accident. And if Bob cannot save this child as an agent, surely AGENT is under-inclusive. We certainly wouldn't want to say that Bob can only be morally required to perform actions he sees reason to perform, given his wacko subjective reasons.

But I think this verdict is not as implausible as might first be thought. First, Bob's status *as* an individual who has any deliberative control over thought and action is severely compromised given his indoctrination. One might think, even in Bob's case, that moral categories are somewhat less than appropriate given his utter failure as a deliberative agent. In this way, we might compare Bob to Joey. Of course, like Joey, morality will require Bob to perform the morally best of those actions Bob can perform as an agent. Nothing about AGENT requires us to assign the moral valence of morally requirable actions for Bob on the basis of Bob's chilling normative beliefs.

Of course, we certainly want to say that there is something very morally wrong with Satanists who refuse to save children, even if their beliefs are so entrenched that they cannot do so as agents. But there is a very important sense in which we can accept this verdict even if we accept AGENT. We would want to say that Satanists behaved badly, for instance, if they *developed* Satanic beliefs—beliefs they could have refrained from developing as agents. We can rightly criticize them for developing their Satanic principles had they the deliberative option to choose other principles, principles the choice of which on their parts would have led to better consequences, or a greater achievement of whatever other moral factor the substantive theory in question deems relevant. Furthermore, if it is within Bob's deliberative control to develop *new* attitudes, he seems morally required to do so. But assume that the fact that these individuals are Satanists is simply beyond their deliberative control. I find it difficult to criticize such individuals morally speaking, so long as they perform actions that are morally best among those they can actually direct themselves to perform as a result of deliberative action rather than luck or mere happenstance.

Of course, we can still make a number of plausible claims about Bob's

refusal. We can say it is bad, regrettable, that Bob did not save the children in question. We will certainly seek to retrain Bob or to get him to change his normative beliefs. We will even *despise* Bob, or seek to *blame* him. It may be that any or all of the *reactive attitudes* are appropriate to direct toward Bob. But I see little ground to be gained in saying that Bob is morally required to perform an action he could not have performed as a result of his own deliberation.

I humbly submit that these are my intuitive reactions, but I am certainly not convinced that they will be widely shared. It may very well be that I have uncovered a seriously implausible verdict of AGENT. But the mere fact that some results of AGENT may be counterintuitive does not entail that we should reject AGENT forthwith. Rather, this consequence should be weighed against the implausible verdicts of competitor theories of moral requirability. In particular, one might wonder whether it is worse or better to accept a theory of moral requirability that entails that on virtually any plausible substantive moral theory, one *almost never* acts as one is morally required to, rather than accepting a theory of moral requirability that entails that individuals are only morally required to act in the morally best ways they could have acted as agents. For my money, even if the latter verdict is implausible in the case of Bob or anyone else, it is far less implausible than simply accepting a view on which we are required to perform actions we will not, given human cognitive limitations, perform. And for this reason, it seems to me, we should accept AGENT rather than TRAD, warts and all.

7. Conclusion

The objection from procedural ignorance is important, and infects virtually all plausible moral theories. I have argued that a fix for this problem is to adopt an alternative to the traditional theory of moral requirability: to insist that x can be morally required to ϕ if and only if x can ϕ as an agent. This avoids the problem of procedural ignorance. Furthermore, AGENT permits of a compelling rationale. AGENT plausibly captures an intuition that some acts—those that could be performed only as a result of happenstance, luck, or instinct—are beyond moral requirability. Also, AGENT is supported by the general thought that there is an important connection between the appropriateness of a moral requirement for a particular agent to ϕ , and that agent's having deliberative control over whether to ϕ rather than perform some other action, as illustrated by Joey's case. Finally, and perhaps most importantly, AGENT adequately avoids the most serious problems that result from morally required action in light of human

cognitive limitations.

Of course, the points in favor of AGENT must be weighed against the points against. Such points there are. However, we must be careful not to jettison AGENT too quickly. First, it is likely that many of our considered judgments, especially considered judgments about just what sorts of reasons we have, are very heavily influenced by the tacit assumption of TRAD rather than a more restrictive theory of moral requirability. Second, and more importantly, on virtually every moral theory, TRAD allows actions to be morally required that we do not believe are, in fact, morally requirable. In weighing the plusses and minuses, AGENT must be taken very seriously indeed.