What Would Lewis Do?

Daniel Nolan, University of Notre Dame

To appear in Beebee, H. and Fisher, A.R.J. (eds). *Perspectives on the Philosophy of David K. Lewis*. Oxford: Oxford University Press.

Penultimate draft 5 March 2021. Please cite the published version when available.

1. Introduction

As well as his better-known contributions to philosophy in areas like metaphysics, philosophy of language, and philosophy of mind, David Lewis says a lot about ethical questions across a range of papers. (Many of them are collected in Lewis 2000.) While individual papers of Lewis's on ethical and metaethical matters have been influential, few philosophers seem to have been influenced by these views as a whole: I know of nobody who describes themselves as a Lewisian in ethics, for example (and if anyone does, they have not said so anywhere Google can find them).

Lewis did not aim to give a systematic moral theory. Nor did he seem to end up with a systematic moral theory by accident, unlike some of the systematic philosophical theorising he ended up doing elsewhere (See Lewis 1983 ix–xi). Still, if we should not hope for a complete moral theory from him, threads of thinking through different contributions can be brought together fruitfully. In this paper, I want to focus most on the interplay between Lewis's theories of rationality and the metaphysics of morals, on the one hand, and his ethical views, on the other.

The first core tension I want to explore is this. Lewis is explicitly anti-consequentialist: "consequentialism is all wrong as everyday ethics" (Lewis 1984 p 214). On the other hand, his theory of rational action is a fairly standard causal decision theory, according to which agents are rational insofar as they maximise expected value. (His preferred variety of causal decision theory is spelled out in Lewis 1981.) It is easy to see how a Lewisian rational agent could behave in line with consequentialist ideals: she would just have to maximise the right sorts of values. But it is harder to see how a Lewisian rational agent could non-accidentally be doing the right thing if the right thing involves a range of constraints other than promoting the right values.

The question of whether Lewis's style of meta-ethical approach can be squared with non-consequentialist first-order views has received much less attention, but I think it is also important. Lewis's meta-ethical position was, in his own words, "subjectivism with bells and whistles" (Lewis 1993 p 105 ftnt 6): in broad outlines, what a person correctly claims is good and right is a matter of that person's attitudes. One advantage meta-ethical subjectivism is often thought to have is that it has a more straightforward story about moral motivation than many rivals: what someone correctly judges to be right has an internal connection to their motives to act. But the natural story for subjectivists to tell about moral motivation works best to explain a tie between moral motivation and moral *value*: explaining motivations to conform to moral constraints that are not a matter of pursuing values is not yet explained by explaining the motivation-moral value connection.

These conflicts are interesting for much more than reasons of intellectual biography. The question of how to marry a theory of rational decision making to non-consequentialist moral theorising is an important one on the current philosophical landscape, with some arguing that it poses a very serious problem for non-consequentialist views (e.g. Jackson and Smith 2006), and others more sanguine about non-consequentialist prospects (e.g. Colyvan et al 2010). While something like standard decision theory (evidential or causal) is widely subscribed to at least as a useful model of rational decision, most philosophers do not accept consequentialism in ethics. (The PhilPapers survey carried out in 2009 reported that under 24% of their target sample of professional philosophers accepted, or leaned towards, consequentialism (Bourget and Chalmers 2014 p 476).) So many philosophers, however sympathetic or not with Lewis's other views, seem tempted to the combination of standard decision theory and anti-consequentialist ethics. Several subjectivist approaches to metaethics put value at the centre of their metaethical account: a simple example would be to say that that what is good is what one desires in ideal circumstances. Subjectivist accounts of ethics that focus on moral value might also need to be adjusted for an anti-consequentialist.

Examining the interaction of the specific ethical and meta-ethical commitments Lewis had seems to me a fruitful way to start exploring the options both for anti-consequentialists who wish to

retain something like standard decision theory, and for a wider class of subjectivist anti-consequentialists than Lewis's particular theory.

2. Morality and Rationality

Lewis's theory of rational decision making requires rational agents to maximise expected value given their beliefs.[1] (They should do so by the canons of causal decision theory rather than evidential decision theory: but this wrinkle will not make a difference in this context.) Presumably it is possible for a moral person to be practically rational: at least, Lewis nowhere suggests that you cannot be both a good person and a rational one, and when he discusses what the right thing to do is in various situations he talks as if people should weigh their credences and values attached to outcomes in roughly the usual way. (For example, when he discusses nuclear deterrence and the moral and practical questions of what to do should deterrence seem to fail in e.g. Lewis 1984 and Lewis 1989b.)

On the other hand, Lewis rejected consequentialism in ethics. As well as Lewis's claim that "consequentialism is all wrong as everyday ethics" (Lewis 1984 p 214), Lewis objects to all "universalistic" ethical systems, including utilitarianism but also, from context, any system that determines right and wrong in terms of agent-neutral moral values, saying "[a]n ethics of our own world is quite universalistic enough. Indeed, I dare say it is already far too universalistic; it is a betrayal of our particular affections." (Lewis 1986 p 128). Instead, "[f]or those of us who think of morality in terms of virtue and honour, desert and respect and esteem, loyalties and affections and solidarity, the other-worldly evils should not seem even momentarily relevant to morality. Of course our moral aims are egocentric. And likewise all the more for those who think of morality in terms of rules, rights and duties; or as obedience to the will of God." (p 127). Lewis here rejects consequentialism in favour of something more like virtue ethics, or an ethics

[1] I do not mean to suggest, of course, that the *content* of a rational agent's desires is that they maximise value, or that they need to conceive of themselves as choosing higher expected value options over lower ones in choice situations: it is rather than their rational activity conforms to descriptions in terms of decision theory in the right sort of way. (Exactly what *is* required for the right sort of conformity is philosophically controversial. Presumably it requires more than that a person in fact behaves in the same way a rational person would, or could.)

that puts at its centre not just facts about virtue but facts about honour, loyalties, solidarity, and so on.[2]

So, on the face of it, Lewis rejected the doctrine that the right thing to do is to maximise the good, or even to maximise the expected good. So on the one hand, rational people are always acting so as maximise a certain quantity ("value", in one sense of that term). On the other hand, moral people do not always act to maximise the good, and presumably this option is not even always permissible for moral people. But plausibly one can be clear-headedly and rationally a good person. So rational and good people are maximising a quantity in their actions, but that quantity is something other than moral value (or some weighted sum of moral values). This is not an inconsistent view, but it is on the face of it strange.

Let me try to articulate why I find that combination strange. If we have a conception of rationality as maximising a function that is a weighted balance of our various goals, the most natural thing to think is that a rational person who is also moral includes moral values among the goals being weighted, and gives them a suitably high priority. That is not to say that even an ideally moral person would have no other components of the eventual weighting: even a saint is allowed to prefer strawberry icecream to chocolate, given other considerations are equal. For all we have said so far, different ideally moral people might have very different desires, provided they serve ideal moral ends ideally well. (Perhaps the director of a global food relief program and an artist producing uplifting and educational art treasures might have different desires about how to spend their day, without either being more morally ideal than the other.)

It seems natural, then, given the picture of rationality, to build in morality to the moral-and-rational agent by building having the agent's overall evaluation of outcomes be largely (but perhaps not entirely) a product of the moral evaluations of outcomes. It is easy to see how these may mesh together for a consequentialist. A maximising act consequentialist (who thinks only the best, or equal-best, option for action in any situation is morally permissible) can say the ideally-moral-cum-rational person has "utilities" that rank the morally most valuable option as

[2] One interpretative note so as not to mislead: in speaking of "egocentric" values, Lewis does not just intend "self-centred" values in the usual sense. Concern for *my* family or *my* country or *my* planet and its inhabitants would all count as "egocentric" in the broad sense Lewis intends here.

their most valuable option, when there is a unique moral best, and when there are several moral options tied for most valuable, their "utilities" rank some or all of those options as the highest. Other consequentialist approaches might vary the link between moral value and the utilities of the ideally moral person in minor ways: some satisficing consequentialists may wish to say that even ideally moral people sometimes rate a morally sub-optimal option as having the most utility, provided it is morally good enough, for example. And any theorist may wish to loosen the restrictions a little when we are dealing with a moral-and-rational agent who is less than ideally moral, or less than ideally rational, even if that agent is very moral and very rational.

This way of meshing moral value and the utilities of a rational agent does not look available to the anti-consequentialist. According to a committed anti-consequentialist, for example, sometimes the option that brings about the most moral value is forbidden. Moral values may not be entirely orthogonal to the agent's utility function, but whatever the connection is, the agent must, if she is to be moral, sometimes assign more utility to options that are valued sub-optimally by the appropriate aggregation of moral values. Indeed, to take Lewis's anti-consequentialist remarks at face value, he seems to be suggesting that production of moral valuable outcomes has relatively little to do with the rational motivations of a moral person. Instead the moral and rational person would act in ways that express virtue, display honour and friendship, etc., even when this came at the cost of the most morally valuable outcomes. This seems to pull the agent's utilities a long way from the moral values of outcomes. So the two ways of thinking about agents are hard to put back together again: rational action is a matter of choosing from what is evaluated as the best, but a moral and rational person is largely insensitive to what is morally best, and may even be (morally) required to act against what is morally best. It seems the moral and rational agent must evaluate outcomes using moral criteria, but largely not by using the moral values, which seem to be the moral criteria best suited for evaluating outcomes.

When Lewis says, for example, that consequentialism is "all wrong as everyday ethics" (Lewis 1994 p 214) Lewis appears committed to the view that consequentialism is often incorrect in its verdicts about what to do, rather than just incomplete or incorrect in its claims about which moral truths are explained by which. He never explicitly claims that we are sometimes forbidden from doing what maximises value, to my knowledge, but one place he comes close is in

discussing an argument that we ought to beg, borrow or steal as much as we can to donate to alleviating poverty in the third world, and indeed that we must devote all of our resources beyond a small amount for our own subsistence to that cause (Lewis 1996b). The context is a discussion of what Unger 1995 seems to be committed to in his discussion of what is meant to follow from "our ordinary morality". Lewis thinks this extreme consequence does not follow from our own moral commitments, but he does indicate that it may follow from a "contentious system of utilitarian ethics", in what is presumably a reference to Peter Singer's views, among others. Lewis appears to endorse the view that we may, permissibly, do less good than the most we can do when that good would be for people suitably "separated" from us (pp 157–158) since he suggests the permissibility of not making great sacrifices for distant strangers is not just part of our "ordinary morality", but is a "legitimate" part of our ordinary morality. And he seems to suggest that stealing and other "unscrupulous" behaviour would not be right even if it maximised the good we could do for those in dire need, at least when those needy are suitably separated from us. I would guess that e.g. stealing so much from one's friends and family that they ended up homeless and destitute, just in order to get some additional money to an effective charity, would be something Lewis would condemn as morally wrong even if the good done for the charity recipients outweighed the harm done by the betrayal and poverty inflicted.

One option for resolving this tension, as suggested above, is to hold that one cannot be both entirely moral and entirely rational at the same time. If that is the way out, we need not understand the moral person as maximising value, of any sort, since that cognitive economy is only required for rational agents. The position that morality and practical rationality are in tension is not a *prima facie* absurd one, despite the fact that many contemporary philosophers see moral requirements as a subset of broadly rational requirements. Kierkegaard argued that religious faith and rationality were in tension, and a similar line of thinking might suggest that a sufficiently whole-hearted embrace of the moral law sweeps us away from our other commitments, even commitments to rationality. Or we might fear, with Sidgwick, that practical rationality counsels a self-interest that can conflict with moral requirements (Sidgwick 1907 pp 496–509). Nevertheless, I will not pursue this option for understanding the apparent tension between Lewis's theory of rational action and his theory of moral action, and I will assume that being moral is normally consistent with being entirely practically rational. Exegetically, I see no

evidence that Lewis saw an automatic tension between being practically rational and being moral; and the approach he takes in discussing moral matters suggests to me a picture of agents deliberating about how to achieve appropriate ends, and what they might best to do accomplish them.

Another way out, that Lewis may well have taken seriously, is that it is just impossible to be an altogether moral person in most realistic situations we find ourselves in. If so, *a fortiori* it would be impossible to be both entirely moral and entirely rational in such circumstances. Lewis took seriously the thought that there were genuine moral dilemmas with no overall right thing to do. Indeed, he went further, arguing that there may well be "dilemmas of virtue", where a decision situation may yield conflicting demands on action arising from a single virtue (Lewis 1998b). Lewis also seemed sympathetic to incommensurability of moral considerations, including questions of what to do (see Lewis 1984 p 207 for incommensurability in assessment of character.) He thought that even an omnipotent being would face dilemmas of value that would not give it a way to be perfectly good in its responses (Lewis 1993 p 103). It would be no surprise if a perfectly rational agent could not be entirely moral in realistic situations, if *nobody* could, in principle, be entirely moral in realistic situations. (That is, they could not be entirely moral because there is nothing they could do that did not violate a moral obligation, or nothing they could do could be entirely moral because it was indeterminate what counted as doing the right thing.)

While important in Lewis's ethical thinking, my suspicion is that we can put aside these dilemmas and incommensurabilities for current purposes. So I will primarily focus on cases where Lewis holds that there is a right course of action, and there is a unique and determinate ranking of the moral value, impersonal and personal, of the options an agent faces. Such cases might be artificially simple compared to many of the complex moral problems we face, but it will make for a more tractable discussion of the apparent clash between Lewis's anti-consequentialism and his decision theory. What we learn from these simpler cases can then be carried over, with suitable adjustments, to more complex ones. (Omnipotent beings cannot avoid these clashes of values, but I take it this reflects an omnipotent being's staggering range of

options rather than it being impossible in general to be faced with a decision where one option is best supported by moral considerations.)

3. What Did Lewis Reject When He Rejected Consequentialism?

We have just seen that Lewis appears to reject several parts of the standard consequentialist package: he says explicitly that it "[c]onsequentialism is all wrong as everyday ethics", though he does go on say it is "right as a limiting case", by which I take it he means it gets the right result in treating duty and honour as less morally important than the potential destruction of thousands or millions of innocent lives (Lewis 1984 p 214). Beyond that, as noted above he condemns any "universalistic ethics" as "a betrayal of our particular affections" (Lewis 1986 p 128). (Though here I suspect he means to condemn any ethical system that is entirely about e.g. the welfare of everyone, rather than a system that contains some universal aspects, while also making room for a special status for our personal commitments and relationships.) Finally, as mentioned above at one point he appears to deny that maximising moral value is even permissible, when it is a matter of helping far-away strangers at the cost of cheating, stealing and betrayal (Lewis 1996b p 127–8).

Let us unpack further what it is about consequentialism that provoked Lewis's rejection of it. While Lewis stated his opposition to consequentialism more often than he argued for that opposition, he did often say some things that indicated what he disagreed with. Reconstructing Lewis's reasoning is unfortunately sometimes speculative given the nature of his remarks, but I will nevertheless make the attempt. One crucial thing will be to unpack what Lewis was attributing to consequentialist views. "Consequentialism" is a term that is used with varying meanings in the literature, and the more we pack into it, the more options Lewis has to disagree with it while not having to take issue with things that mesh well with his picture of rational decision making. While I do not know exactly what Lewis intended by the expression, let me discuss a few doctrines that authors sometimes include as part of the family of views they identify as consequentialist.

The core of consequentialism, as I understand it, is the doctrine that the key commitment of an ethical theory is an account of the moral goodness, or moral value, of outcomes. A theory of other moral matters such as the rightness and wrongness of actions, or the virtuousness or viciousness of character traits, is then to be given derivatively in terms of moral value. A simple version of consequentialism is classical Benthamite utilitarianism, where the balance of pleasure over pain is the only good; where every person at every time has some amount of utility defined in terms of that person's pleasures and pains; every person's utility is as important as any others; and that right action consists in acting so as to maximise overall utility: that is, to maximise the sum of the utilities of all agents, compared to the sum of utility of all agents resulting from alternative actions.

Consequentialists need not adopt the utilitarian account of what moral value is; nor do they need to define the total moral value of an outcome by an unweighted sum of the moral goods and evils accruing to each agent. Outcomes not even involving agents might get some positive or negative value, for example, or we might weight agents' values when calculating an overall sum—to privilege the worst off, or to privilege the aristocrats, for example. We may wish to use *expected* value in our moral theorising instead of what the value of outcomes will or would in fact be, and we may wish to impose much less structure than standard measure theory on comparisons between values of outcomes. Some authors (Scheffler 1982 p 1, Pettit 1997 p 129) have wishes to insist, in addition, that a consequentialist theory must take moral value to be *agent neutral*. That is, the moral good is the same for all agents. On the other hand, others are willing to count "agent relative consequentialism" as a species of consequentialism. This would allow that different agents have different moral goods, and different agents appropriately aim to maximise (or satisfice) the good associated with that agent. (For a recent overview, see Hammerton 2020.)

I use the terms in such a way that so-called "agent relative consequentialism" counts as genuinely consequentialist, and I take it that its theoretical similarities to agent-neutral consequentialism justifies discussing the two together in many discussions of consequentialism. I am also inclined to think that whether we use the label "consequentialism" broadly enough to cover agent-relative consequentialism is primarily a terminological matter, and beyond considerations of taxonomic

usefulness there is not a further deep question about whether agent relative consequentialist proposals are "really" consequentialist or not.

This terminological issue is worth flagging here, however, since it is not clear to me whether *Lewis* was using the expression "consequentialism" broadly enough to cover agent-relative consequentialism. If he was thinking that consequentialism was by definition agent neutral, and his intuitions were agent relative, that might be a partial explanation of his rejection of "consequentialism", as he understood it. Some evidence that he may have taken "consequentialism" in his sense to require agent neutrality of moral value will be discussed in section 6.

Simple forms of consequentialism offer a unified account of moral value. But it is also possible to be a consequentialist and a pluralist about moral values, as e.g. G.E. Moore was. Relatively uncomplicated forms of pluralism do not change the overall picture much: when several goods are entirely commensurable, we can calculate an overall utility for each and give a final score to each outcome. Things get trickier, though, when different values are incommensurable, as Lewis thought they could be. Perhaps the most explicit statement of this is in Lewis 1993, where he states his "own conviction that values are diverse and incommensurable and in conflict in ways that even God could not pursue some without betraying others" (p 103).

Lewis at least endorsed the agent-relativity of some moral values; he was a pluralist about moral value; and as mentioned in section 2 he thought there was a good deal of incommensurability and even dilemmas between moral considerations. He appears to have thought that approaching morality as a matter of maximising value was wrong-headed, and that it was not always permissible to maximise moral value, though we may eventually need to explain away those appearances, as I will discuss in section 6. For now, then, let me turn to one way someone who accepts typical credence-and-utilities decision theory might try to accommodate moral considerations and constraints that do not flow from moral values and attitudes to moral values.

4. Maximising Quantities and a Euthyphro Question

The mere fact that an agent can be represented, in her actions so far as they are rational, as maximising a quantity given their beliefs, is not enough to show that the underlying *explanation* of those actions, or the rationality of those actions, is a matter of maximising a quantity given the agent's beliefs. In principle, a vast array of dispositions to act can be associated with a function maximising a quantity, and there need be no interesting link between that representation and the actual psychological life of the person exhibiting that behaviour. Indeed, a very wide range of patterns of obligation, permission and forbidding can be modelled as requiring an agent to maximise a defined-up quantity in the agent's action (Oddie and Milne 1991). That does not make every moral theory consequentialism in disguise, albeit with arbitrarily alien-looking conceptions of value, nor does it make almost every theory of rational permission into a desire-satisfaction one, albeit with arbitrarily alien-looking contents of desire.

The philosophical picture behind Lewis's theory of rational decision is more than that rational agents can be represented so that their actions correspond to certain relationships between their credences and utilities. Lewis endorsed a broadly Humean picture of motivation, according to which the degrees of belief and utilities associated with an agent's beliefs and desires play a central role in the psychological explanation of the agent's actions insofar as she counts as an agent at all. Still, there is room even for Lewis to draw a distinction between an agent maximising expected consequences, and the explanation of the agent's action being in terms of maximising that quantity. In particular, a Lewisian rational agent can act to maximise a given quantity *because* of what she ought to do, rather than something being what she ought to do because of the structure of her actual or idealised credence and utility functions. Or so I will argue. The general idea will be that an anti-consequentialist can see the moral agent as recognising obligations or recommendations of virtue, and because of recognising those facts, come to have desires and utilities she would otherwise lack, that in turn go together with the agent's beliefs to both cause and rationalise action.

To warm us up, consider an ideally rational Lewisian agent following orders. The Captain orders Agent to climb the hill and plant a flag. Agent desires to follow orders, and believes with very high credence that her orders are to climb the hill and plant a flag, on the basis of hearing Captain. So, maximising expected desire satisfaction (given the absence of other strong desires),

Agent tramps to the top of the hill and sets up the flag. Even though the agent maximised an expected value in her action, the order and her obligation is not downstream of her utility function in explaining why she acted as she did. Rather, she had the relevant pair of utility and credence functions that led to her action *because* of the order and the accompanying obligation. (My own view is that we can model this agent either as changing her credence and utility functions after the order, or only her credence function after the order. But either way we do it, it should be clear that the order and obligation are not downstream of her credences plus utilities.)

An anti-consequentialist, who rejects an account of moral obligation as being determined by moral value of outcomes, could tell a structurally similar story about how an agent acts on a moral obligation. Agent hears Duty, "stern daughter of the voice of God", tell her that she has a moral obligation to donate $100 to charity. Agent wants to follow the dictates of duty, and believes with high degree of belief the dictate is to donate $100 to charity. Acting on her updated credence-utility pair, Agent donates $100 to charity. Agent's action is proximally explained by the interaction of her credences and utilities, but her being obliged to donate $100 is not explained by those functions, and perhaps not explained by any value function. Still, the obligation can play a role in explaining why she donated the $100 to charity: she has the credence and utility functions she has partly because of the obligation, and in this case we can trace the explanation of the action back to duty's orders.

This initial sketch is crude in several ways. I gather deontological constraints are not, in fact, whispered into deontologists' minds by personified Duty. More seriously, perhaps we are not supposed to become aware of our duties by changing our credences first. Nor, perhaps, are our moral motivations supposed to be summed up by a general desire to carry out our duties, whatever they are. Even if we tinker with these features of the initial sketch, we can still end up with a story where duties explain credences and/or utilities, in ways that thereby explain the actions of a moral and rational person, rather than duties being explained by value facts, whether facts about the agent's own utilities or some more objective facts about moral value. (I will resist the temptation to sketch several ways we might put some flesh on these bones.)

Provided we had some story about how a good person came to opinions about duty, virtues and the rest; and we had some story about how those beliefs interacted with her motivational set, either in terms of de dicto desires to do the right thing or in some more indirect way that did not go via her utilities already being aligned with moral values of outcomes; then we would have the material to explain the connection between a moral agent's rational systems and the moral truth, or at least what she takes the moral truth to be. (We may wish to add a story about how her evidence appropriately bears on her moral beliefs; or some constitution of obligations and virtues for her out of her attitudes; or to otherwise explain why her moral beliefs and the moral truth can do better than align merely by chance.) She will be acting, insofar as she is rational, to maximise a quantity, but that quantity may have little to do with what is morally valuable: in the same way that our soldier planting a flag on the hill does in the case above not do so because she antecedently valued flags on hills, but because she rationally follows orders. We may even want to account for moral values themselves in terms of what well informed moral agents bring about, should we choose to: the direction of dependence between moral value and e.g. obligation would still be different from the consequentialist's preferred explanation of right action in terms of moral value.

An additional wrinkle in making sense of Lewis's view is that he does think consequences play some role: and indeed when the stakes are high enough, consequences trump other considerations. (Lewis 1994 p 214) This must mean that there is some way for consequences to interact with other moral considerations, so that those other considerations can make more of a difference in some cases than the balance of consequences, but the non-consequentialist drivers of action can be somehow outweighed in the good agent when the stakes are high enough. The most natural way to model this would be to assign all the different considerations weights and do an expected utility calculation: if we started off thinking the "consequences" were a matter of life, liberty and happiness, for example, we could add a range of agent-relative weights for the agent not breaking promises, the agent doing her duty, the agent behaving virtuously herself etc. Assigned the right weights, when the life-liberty-happiness stakes are low, the agent focuses on avoiding breaking promises, doing her duty etc., while when the life-liberty-pursuit stakes get high, the agent lies, cheats and kills for the greater good. If a Lewisian proceeded this way, they would need to show how to understand this formalism without covertly treating duty-doing,

promise-keeping, virtue-expressing etc. as fundamentally matters of moral value after all. Presumably this could be approached in the same way as above, treating the utilities assigned to these matters explained by obligations and virtues, rather than explaining either in terms of values.[3]

5. A Second Route to Consequentialism for Lewis

As far as maintaining the combination of causal decision theory and non-consequentialist ethics, the options sketched in the previous two sections may be sufficient. However, there is one more feature of Lewis's overall account that may commit him to consequentialism, and here I think the prospects for avoiding an ultimate embrace of consequentialist ethics are less rosy than the prospects of resolving the tension we have been looking at so far.

As mentioned above, Lewis was willing to describe his own account of moral as "subjectivism with bells and whistles" (Lewis 1993 p 105 ftnt 6). The main place where Lewis explores this view is in an account of value he provides in Lewis 1989a. This paper is about what values are in general, though as I read it Lewis thinks of moral values as a subset of the values he discusses. (He treats "to be a value" as amounting to "to be good, near enough" when discussing the connection between his view and the subjectivism about moral value targeted by G.E. Moore on p 71. I take it the "near enough" is because he wants to include values we may not think of as moral values in his account.)

On Lewis's account, to be a value is to be "something of the appropriate category... [such that] we would be disposed, under ideal conditions, to value it" (p 68), where valuing *A* is a matter of desiring to desire *A* (p 71), and the ideal conditions are certain idealised conditions of imaginative acquaintance (pp 79–83). Who "we" are may depend on how widely these dispositions, under ideal conditions, to desire to desire are shared: in the limit case, someone talking of what is valuable may only be speaking for himself. Lewis says that one of the

[3] There are other ways of attempting to let consequences trump when the stakes are high, but have consequences trumped in ordinary stakes cases: there are various proposals for treating non-consequentialist considerations as "side constraints" on the pursuit of moral value. If these side constraints cut out at some value threshold, various strange results follow: see Alexander 2000. If they do not, however, things seem even less satisfactory: see Nolan 2009.

advantages of this view is that it is "naturalistic", both in the senses that it is (more-or-less) a kind of analytic naturalism, defining the moral in non-moral terms, and in that it "fits into a naturalistic metaphysics. It invokes only such entities and distinctions as we need to believe in anyway, and needs nothing extra before it delivers the values". (p 68) The strong suggestion is that Lewis would like to find a naturalistic account of the truth in ethics in general. (Not to mention Lewis's more general commitment to materialism and naturalism: see Nolan 2005 chapter 2.)

One of the reasons to adopt this sort of account of moral value is that it provides an easily intelligible link between moral value and moral motivation. It is a widespread thought that moral value should somehow be *practical*: at least in typical cases, we are motivated to pursue the moral good (though of course that motivation may be less strong than others; and we may make moral mistakes or otherwise screw up). If what is morally valuable for us is what we desire to desire, then even if we do not desire the good, we at least want ourselves to: and when our desires line up with our desires to desire, we will be motivated to pursue what is morally valuable for us. (Since Lewis does not hold that desiring to desire is enough for valuing correctly, but requires desiring to desire as we would be disposed to do were we ideally acquainted with the putative value, the connection he offers between what is valuable and what we desire is even more indirect.) Still, in circumstances relatively like the ideal, at least people with a certain kind of coherence in their desires will be disposed to pursue morally valuable outcomes.

There is a limitation for this line of thought, however, for a Lewisian subjectivist who rejects consequentialism. Someone whose desires line up with what they desire to desire (when they manifest the disposition to desire to desire that they would have in ideal situations of acquaintance) will be someone who desires what is morally valuable for them. And so, given that desires are motivations, they will be motivated to pursue the moral good (or their moral good at least, given the background subjectivism). Non-consequentialists, however, typically hold that there are moral requirements other than pursuing the moral good, and which may even forbid choosing more moral good over less in some circumstances. The problem is to give an account of the *motivational force* of these non-consequentialist requirements, against the background of

Lewis's brand of subjectivism. Why wouldn't the good, practically coherent person, who desires what she desires (etc.), just ignore those requirements when it gets in the way of pursuing moral value? That is what they want to want to do, when in suitably ideal conditions.

Lewis has several options open to him to resolve this problem. He could endorse the conclusion that his story about moral value has nothing to say about the motivation of non-consequentialist requirements, and even decline to put anything in their place: after all, something being the morally right thing to do is one issue, and whether any of us are non-accidentally motivated to do the right thing is another. Giving entirely separate motivational stories about different moral obligations is uncomfortably disunified, and in any case, Lewis is likely to find the motivation to obey moral requirements somewhere in our desires, for two reasons. One is his general belief-desire psychology for human action (his "Humeanism" about motivation). Another is his argument that a range of anti-Humean suggestions about moral motivation in particular face fatal objections: his "triviality results" for "Desire as Belief" theses. (Lewis 1988a and 1996a). So a Lewisian should not be setting up duties or virtues as things that have a motivational grip on us except via a connection to our desires.

The moves available to Lewis in section 4 do not immediately help here: even if our desires could be explained by our responding to facts about obligations or facts about virtues, that would not go very far to explain where these non-evaluative aspects of the world come from, and how they can be otherwise captured in naturalistically respectable terms.

Space precludes further speculation about how a resolutely anti-consequentialist Lewisian might try to find a naturalistic place for moral phenomena that are not to be explained in terms of moral values. As well as providing a naturalistically acceptable account of actions flowing from virtues; duties and obligations, together with a motivational story about all of these, such an account would need to explain how a morally good person is to balance these considerations in action. Lewis declined to offer an account even of balancing values in Lewis 1989, though he did think that some moral considerations could sometimes outweigh others (see e.g. Lewis 1984 p 214). This story might yield a lot of incommensurability and perhaps even dilemmas, if pursued in an entirely Lewisian spirit. The resulting story might not be as systematic as many extant

proposals, but the tenor of Lewis's moral remarks suggests that he would have seen this messiness as true to our "common sense morality" and our natural moral opinions, rather than something to be swept away by philosophical theory-building.

6. New Evidence About Lewis's Own View

Above I suggested that adopting agent-relative consequentialism is one way out some of these tensions for Lewis. We can recast some of his objections to consequentialism as only objections to *agent neutral* consequentialism, and then reconstruct his view as placing value in the centre of the story both of a theory of action for moral agents, and his subjectivist story about the moral truth. There are more thorough-going anti-consequentialist options too, explaining rational action partly in terms of moral phenomena not to be explained in terms of moral values, and treating Lewis's meta-ethical picture of values as only part of the needed naturalistic picture of ethics.

While Lewis has at least these two options open given his published work, we can make some interpretive progress by looking two letters David Lewis wrote to Philip Pettit in 1991 about consequentialism and non-consequentialism, published as letters 707 and 708 in Beebee and Fisher 2020 (pp 494–498). The best way of interpreting Lewis here seems to me that he was, after all, an agent-relative consequentialist, though it is hard to be sure. Lewis is responding to a verbal discussion with Pettit, a letter from Pettit, and an offprint Pettit sent him as well. The offprint was of Pettit 1991 (Beebee and Fisher 2020 p 497 ftnt 1).

Lewis's letters related to Pettit's honouring/promoting distinction for treating moral values, a distinction Pettit took to characterise the difference between consequentialist and deontological approaches to values. Pettit claims that both consequentialists and deontologists have an important place for moral values in their systems. The difference, as Pettit sees it, turns on how the two approaches treat the moral values they hold dear. Pettit draws a distinction between *promoting* a value and *honoring* that value. Promoting a value is familiar: to promote a value X is to bring about outcomes that score more highly on the ranking function associated with X. (Promoting happiness is bringing about more happiness overall.) Honoring a value is different. While I wish to focus on Lewis's rival to Pettit's distinction, Pettit himself holds there are a variety of attitudes we can take to values besides promoting them: we may seek to exemplify

them ourselves with relatively little attention to whether others do so; we may try to keep clean hands with respect to them (e.g. honouring loyalty by not engaging in betrayal, whatever the overall consequences for loyalty in the world); or we may even act in ways that are not consciously sensitive to the value, though we may be evaluated positively by that standard by others. (One way to be loving is to not care about love per se, but rather to care about the beloved.) See Pettit 1989 p 119–120, Pettit 1991 p 19.

Lewis points out that there is no agreed definition of consequentialism vs deontology, so attempts to characterise the distinction should not be to "capture a pre-existing sharp distinction", but rather the objective should be to "put a finger on one of the differences" between prototypical consequentialists and proto-typical deontologists. (Beebee and Fischer 2020 p 495). That said, he offers three characterisations of approaches to values that together cover the ground that Pettit's distinction is intended to cover. The first is "pursuing a value by trying to instantiate it, where V is a value you can instantiate, namely a value *de se*" (p 497). That is, for paradigm cases, trying to ensure that you yourself have the property, or have it to a high degree. Valuing courage in this way is connected to desiring that oneself be courageous, valuing being a good friend in this way is valuing oneself being a good friend, and so on. In typical cases, this is similar to the first way Pettit talks of honoring a value. Lewis goes on to call this "pursuing" a value, presumably because of the connotation of pursuing it for oneself rather than "promoting", which sounds more general.

The second way, "promoting" a value, comes in two varieties. (2A) is promoting V by trying to make it widely instantiated in one's world. This is equivalent, for Lewis, to trying to ensure one is an inhabitant of a world where V is widely instantiated, since Lewis thinks we can construe every desire as a desire *de se*. (See Lewis 1979 and Nolan 2006.) (2B) is to promote U (*sic*) by trying to ensure one's world instantiates U, or instantiates it to a high degree. Again this is equivalent, according to Lewis, to a certain kind of desire *de se*, the desire that one lives in a world that has *U*. What is distinctive about both of these desires is that they can be characterised to be "agent-neutral": we can find a property that everything in a world shares that is being pursued by each person in that world who has that value. (If I value a maximum feasible balance of pleasure over pain in my world, then all my worldmates with the same value are pursuing the

same goal, the maximising of pleasure over pain. While if I value courage in the sense of the previous paragraph, and you do too, then I am trying to ensure I am courageous and you are trying to ensure you are courageous.) 2A and 2B are varieties of promotion of value, in Pettit's sense, even though they also be characterised, according to Lewis, in terms of desire *de se*. The important difference is the kind of property that is being desired *de se*.

Lewis distinguishes a third approach to valuing a property: "'honouring' a value by trying to have clean hands with respect to it", probably in response to Pettit treating it as an alternative to Lewis's (1) and (2A and 2B). Lewis wants to assimilate this to versions of the options he has already considered: "having clean hands with respect to a value V" is just another value, and one we can pursue (option 1) or promote (options 2A and 2B) (Beebee and Fisher 2020 p 498).

Lewis's discussion in these two letters suggests that he thinks these approaches to moral values cover the options, or at least the plausible options. And he suggests that the main relevant distinction for taxonomy is the one between his option (1) and his option (2), since we should assimilate option 3 to one or the other, depending on its flavour. So it is reasonable to infer that if Lewis saw himself as rejecting consequentialist approaches, he should reject the view that morality is centrally about promoting agent-neutral values, which is what consequentialist-style theories enjoin in Lewis's taxonomy. Instead, Lewis must be thinking that plausible non-consequentialist approaches give a central role to "pursuit" of values in Lewis's sense: seeking to instantiate properties oneself that are not automatically shared with all of ones worldmates.[4]

If this is right, then Lewis is in effect endorsing the view that anti-consequentialists, like himself, are adopting *agent relative* moral values, and are pursuing more rather than less of them. On this reading of Lewis's response to Pettit, then, Lewis is more-or-less an agent-relative consequentialist after all, albeit one who makes some place for agent-neutral values; and one who wishes to at least leave open the possibility of incommensurable values and genuine moral dilemmas (p 495). Lewis's rejection of "consequentialism" in other places must then be centered on the agent neutrality of what Lewis took to be the position of the "consequentialist". And we

[4] These properties need not be intrinsic properties: the property of having keeping one's promises typically involves others who are the recipients of those promises, and the property of having loving relationships with one's family members is typically a relational property involving one's family, to give just two examples.

should not forget that this agent-neutrality can revolt the sensibilities of a moral theorist, especially for some kinds of apparently moral value. If you value loyalty as an underived moral value, for example, the suggestion that one might betray all of one's own loyalties to ensure people on the other side of the world are more steadfast in their loyalties, loyalties which you do not share, may seem not just monstrous but absurd. If you morally value your relationships to your family, cutting all your family ties so that unrelated people a hundred years hence can have slightly closer ties with their families is likely to seem repellant: and if you are willing to disrupt your family at the drop of a hat to promote better family relationships for strangers, your own family might reasonably doubt whether they meant much to you in the first place.

One reason to be cautious in drawing the conclusion that, after all, Lewis was an agent-relative consequentialist in anti-consequentialist clothing, is that in his discussion with Pettit Lewis is concerned to draw a distinction he puts in terms of the "consequentialist" and the "deontologist". In places like Lewis 1986 pp 127, above, he appears to *contrast* his own preferred views with views that give pride of place to "rules, rights and duties". While he does not use the expressions "virtue theory" and "deontology" in the 1986 passage, it strongly suggests he prefers an approach more like a virtue-theoretic one to a deontological one. I am interpreting Lewis's distinction in his response, in effect, as capturing what he thinks of as a distinction between archetypal kinds of consequentialist and archetypal kinds of anti-consequentialists, which is also what Pettit is trying to characterise with the honouring/promoting distinction. But one might worry that Lewis thought archetypal virtue theorists required some *third* characterisation, and were characterised neither as neither pursuing agent-neutral moral values nor pursuing agent-relative moral values. In his letters to Pettit he does not suggest a third characterisation, which makes me think that in this context at least the "deontologist" contrasted with the "consequentialist" is supposed to be characterised broadly enough to cover the range of standard anti-consequentialist positions. Another piece of evidence for this reading is that in the letters to Pettit Lewis identifies G.E.M. Anscombe as the prototypical deontologist, while many read her as being closer to a virtue ethicist. (For the interpretive controversy about Anscombe, see Driver 2018 section 5.1.) So there is some evidence Lewis was thinking of a contrast between consequentialists and anti-consequentialists more generally, rather than characterising "deontologists" in a narrow sense. However, without more textual information this reading could be disputed.

Even with this reason for caution in mind, I think the best way to interpret Lewis in light of the letters to Pettit is as an agent-relative consequentialist, albeit one that is suspicious of smooth aggregation across values and open to the existence of moral dilemmas where no option is morally permitted. It makes natural sense of Lewis's conception of us as both (imperfectly) rational agents and as moral agents: insofar as we are rational we seek to maximise utility of a certain sort, it is just that the moral agent's utility function does not rank outcomes too differently from her aggregated values, both "impersonal" values for the world and for agent-relative values for her. It makes sense of the centrality of value to Lewis's metaethical story: once we determine what the moral values are for an agent, behaving well is doing well enough in choosing options scored by those values out of the agent's available options. The job of Lewis's naturalised metaethics would be finished once it explained moral values in descriptive terms, since the values, agent-relative and agent-neutral, could be the fundamental moral truths through which we explain other moral phenomena. It also makes sense of the scattered positive remarks about what Lewis takes to be morally important: loyalty, solidarity, friendship, honour, integrity, avoiding being the cause of evil, etc. These are the sorts of values naturally modelled as agent-relative, though Lewis also seems to give some role to agent-neutral values, such as valuing the lives and happiness even of strangers (albeit perhaps only the strangers who inhabit our possible world).

It does leave an interpretive puzzle: why would Lewis say, e.g. that consequentialism is "all wrong as everyday ethics" if he is a consequentialist with a few tweaks? And why would he suggest that doing the most we can for those most in need is not even permissible, if it comes at sufficient cost to ourselves and our integrity (Lewis 1996b p 154–155)? Part of the explanation may be that he is thinking of consequentialism (in his sense) as recognising only agent-neutral values, while Lewis holds that in everyday moral matters is it agent-relative values that ought to dominate. (We might be maximising expected value when we decline to steal from our first-world friend to send more money to Oxfam, if it is a case of the agent-relative values of friendship and integrity outweighing the agent-neutral value of relieving famine.) Part of the explanation may be that that the archetypal consequentialist embraces only a few fundamental values: happiness, the absence of suffering, desire satisfaction, etc., and even if Lewis is an agent-relative consequentialist, he is a pluralist who emphasises goods not on those lists. Partly it

may be that Lewis thinks there is much more incommensurability of value and conflicts of fundamental values than an archetypal consequentialist might. A final possibility is that Lewis's commitment to an abstract moral theory was less constant than his clear-minded commitments in many other areas of philosophy: perhaps his rejection of standard consequentialism sometimes manifested in his beliefs as a variety of agent-relative consequentialism, and sometimes manifested as an attitude incompatible with consequentialism even in this broad sense.

7. Conclusion

While there is a natural affinity for a decision-theoretic account of rational decisions with a consequentialist approach to ethics, having one if you have the other is far from compulsory. The moral agent can still be rational by the lights of orthodox decision theory (or its causal-decision-theory counterpart) if she desires to do the right thing, or exemplify virtue, as well as desiring outcomes that are directly morally valuable. Nevertheless, the story is cleaner if the moral agent only needs to do well in bringing about morally valuable outcomes, whether morally valuable in an agent-neutral or agent-relative way. (Or needs to *attempt* to do well, if expected value is what is important.)

I have also argued that there is an affinity between the kind of subjectivist account of moral values that Lewis endorsed, on the one hand, and consequentialism, on the other. Holding back from bringing about the outcomes we most morally value, for apparently moral reasons, is hard to understand. What makes sense to do given our moral motivations is naturally understood as maximising with respect to moral value, insofar as we can; or at least to prefer the morally better to the morally worse. Lewis's subjectivism, as stated, leaves little obvious room for deontology or virtue ethics (or at least for a virtue ethics that goes beyond adding to our basket of moral goods some moral goods concerning our own characters).

As far as I can see, Lewis would have had an intelligible and in some ways attractive set of moral commitments if he had adopted a broadly consequentialist theory, albeit one with room for agent-relative value and genuine moral dilemmas. Given the evidence of his letters to Pettit, perhaps that was in fact his view. While Lewis does not seem to have aimed for a systematic set

of moral doctrines, I have tried to bring out puzzle about why his published views appeared feature a central internal tension: motivational and metaethical stories that naturally suggest consequentialism, with frequent rejections of consequentialism when discussing moral issues themselves. The puzzle is in part historical: what did he think the best resolution was, and if he was, in effect, an agent relative consequentialist, why does that come through so unclearly in his published work? It is in part philosophical: what is the best way for someone sympathetic to Lewis's starting points to develop an anti-consequentialist theory, whether or not this coheres with Lewis's preferred reconciliation?

Lewis's partial and scattered commitments, at least in his published writings, offer a challenge to the intellectual biographer trying to work out what Lewis thought, as well as the ethicist interested in whether there is a distinctive and interesting theory to be developed here. But I think the attempt to synthesise something general to say from scattered and sometimes apparently conflicting commitments is the task that many of us face when trying to work out the truth in ethics, either as a philosophical project or just to guide our own decisions and behaviour. Even if the best we can do for Lewis is a patchwork theory, with different kinds of considerations that do not nicely reduce to each other, and which may often give no verdict about an action or too many, then at least it can be said for such a patchwork that it resembles the implicit moral commitments many of us operate with, whatever we say in our philosophical moments.[5]

References

Alexander, L. 2000. "Deontology at the Threshold". *San Diego Law Review* 37 pp 813–912.

Beebee, H. and Fisher A.R.J. (eds.). 2020. *Philosophical Letters of David K. Lewis: Volume 2: Mind, Language, Epistemology*. Oxford: Oxford University Press.

Bourget, D. and Chalmers, D.J. 2014. "What Do Philosophers Believe?" *Philosophical Studies* 170.3 465–500.

Colyvan, M., Cox, D. and Steele, K. 2010. "Modelling the Moral Dimension of Decisions". *Nous* 44: 503–29.

[5] Thanks to Sara Bernstein, John Bigelow, Frank Jackson, Philip Pettit, Wolfgang Schwarz and the audience at the *David Lewis and His Place in Analytic Philosophy Conference*, University of Manchester.

Driver, J. 2018. "Gertrude Elizabeth Margaret Anscombe" in Zalta, E. (ed.) *The Stanford Encyclopedia of Philosophy*, Spring 2018 edition.
<https://plato.stanford.edu/archives/spr2018/entries/anscombe/>
Accessed 21 February 2021.

Hammerton, M. 2020. "Relativized Rankings" in Portmore, D. (ed.), *The Oxford Handbook of Consequentialism*. New York: Oxford University Press, pp 46–66

Jackson, F. and Smith, M. 2006. "Absolutist Moral Theories and Uncertainty". *Journal of Philosophy* 103.6: 267–283

Lewis, D. 1979. "Attitudes De Dicto and De Se". *Philosophical Review* 88.4: 513–43.

Lewis, D. 1981. "Causal Decision Theory". *Australasian Journal of Philosophy* 59: 5–30.

Lewis, D. 1983. *Philosophical Papers, Volume 1*. Oxford: Oxford University Press.

Lewis, D. 1984. "Devil's Bargains and the Real World" in MacLean, D. (ed.) *The Security Gamble: Deterrence in the Nuclear Age.* Totowa NJ: Rowman and Allenheld, pp 141–154. Reprinted in Lewis 2000. Page numbers are for Lewis 2000.

Lewis, D. 1986. *On the Plurality of Worlds*. Oxford: Blackwell.

Lewis, D. 1988a. "Desire as Belief". Mind 97: 323–332. Reprinted in Lewis 2000.

Lewis, D. 1988b. "The Trap's Dilemma". *Australasian Journal of Philosophy*. 66: 220–23. Reprinted in Lewis 2000.

Lewis, D. 1989a. "Dispositional Theories of Value". *Proceedings of the Aristotelian Society, Supplementary Volume* 63: 113–37. Reprinted in Lewis 2000. Page numbers are for Lewis 2000.

Lewis, D. 1989b. "Finite Counterforce" in Shue, H. (ed). *Nuclear Deterrence and Moral Restraint*. Cambridge: Cambridge University Press, pp 51–114.

Lewis, D. 1989c. "Mill and Milquetoast". *Australasian Journal of Philosophy* 67: 152–171. Reprinted in Lewis 2000.

Lewis, D. 1993. "Evil for Freedom's Sake?". *Philosophical Papers* 22: 149–172. Reprinted in Lewis 2000. Page references are to Lewis 2000.

Lewis, D. 1996a. "Desire as Belief II". *Mind* 105: 303-313.

Lewis, D. 1996b. "Illusory Innocence?" *Eureka Street* 5: 35–36. Reprinted in Lewis 2000. Page references are to Lewis 2000.

Lewis, D. 2000. *Papers in Ethics and Social Philosophy*. Cambridge: Cambridge University Press.

Nolan, D. 2005. *David Lewis*. Chesham: Acumen.

Nolan, D. 2006. "Selfless Desires". *Philosophy and Phenomenological Research* 73.3: 665–679.

Nolan, D. 2009. "Consequentialism and Side Constraints". *Journal of Moral Philosophy* 61.2: 5–22.

Oddie, G. and Milne, P. 1991. "Act and Value: Expectation and the Representation of Moral Theories". *Theoria* 57.1–2: 42–76.

Pettit, P. 1989. "Consequentialism and Respect for Persons". *Ethics* 100: 116–26.

Pettit, P. 1991. "Consequentialism" in Singer, P. (ed). *A Companion to Ethics*. Oxford: Blackwell, pp 230–240.

Pettit, P. 1997. "The Consequentialist Perspective" in Baron, M., Pettit, P. and Slote, M. (eds.), *Three Methods in Ethics*. Oxford: Blackwell, 92–174.

Scheffler, S. 1982. *The Rejection of Consequentialism*. Oxford: Clarendon Press.

Sidgwick, H. 1907. *The Methods of Ethics*, 7th edition. London: Macmillan.

Unger, P. 1995. *Living High and Letting Die: Our Illusion of Innocence*. Oxford: Oxford University Press.