

COMMENTS ON RALPH WEDGWOOD'S *The Nature of Normativity*

RICHARD HOLTON, MIT

Ralph Wedgwood has written a big book: not in terms of pages (though there are plenty) but in terms of scope and ambition. Scope, in that he addresses many of the central issues around normativity, providing an account of the semantics of 'ought', and then a metaphysics and an epistemology to go with it; ambition in that so much of this is novel. Along the way there are myriad discussion of relevant philosophical background issues and of methodology.

Amongst such riches, my focus here will of necessity be limited. I shall concentrate on the first part of Wedgwood's book, where he pursues a three part strategy in which he aims to:

- (i) identify a notion of judgment internalism that captures the essence of 'ought';
- (ii) build a semantics on the basis of it;
- (iii) identify a logic of 'ought' on the basis of this semantics.

So there are two crucial transitions, which I shall take in turn: from internalism to the semantics, and from the semantics to the logic. Wedgwood's aim throughout is to produce an account that, whilst not being reductive, since it will use normative vocabulary in the account of the normative, is nonetheless informative. The contrast here is with quietist approaches that deny that any informative account is available.

My aim is to raise two worries. The first, a point of detail, is that Wedgwood's semantics does not quite capture the judgment internalism that he wants it to capture. The second, more substantial, is that the norms governing 'ought', and the norms governing our practical plans are more distinct than Wedgwood's account would have us think.

INTERNALISM AND THE RESULTING SEMANTICS

Wedgwood's basic idea here is that 'ought' is to be understood by means of conceptual role semantics; that is, it is to be understood in terms of the core conceptual transitions that mastery of the term licenses or requires. A grasp of 'and' enables a thinker to move from an acceptance of 'p' and an acceptance of 'q' to an acceptance of 'p and q', and likewise from an acceptance of 'p and q' to an acceptance of 'p' and of 'q'. A familiar idea at least since Gentzen, is that an account of the semantics of conjunction can be provided in terms of rules that correspond to these two transitions: an introduction rule that corresponds to the first, and an elimination rule that corresponds to the second. If these are to provide an account of the semantics of 'and', then it had better be the case that a grasp of them is sufficient for an understanding of it; there cannot be any more that one needs to know.

Wedgwood argues that, in a similar way, a grasp of the concept expressed by 'ought' is to be understood in terms of the core conceptual transitions that it licenses or requires. And the central transition, according to Wedgwood, is that identified by *Normative Judgment Internalism*:

Necessarily, if one is rational, then, if one judges 'I ought to F', one also intends to F. (p.25)

and

Necessarily, if one judges anything of the form 'I ought to F' then one also has a general disposition to intend to do whatever one judges that one ought to do. (p.28)

These claims are clearly controversial, though some of the controversy is avoided when it is realized that the 'ought' in question is supposed to be a general deliberative 'ought', the 'ought' of practical reason, and not the narrow 'ought' of moral obligation: so it is quite possible that one may conclude in certain cases that the thing one ought to do is not the thing that one morally ought to do. (Though, of course, one cost of this is that Wedgwood does not provide us with an account of the specifically moral 'ought'.) Still, I am not going to question the broad thrust of the internalist position at this point. I instead look to the way that Wedgwood develops this internalism into a semantics for 'ought'.

Taking 'ought' to be an operator on propositions that is indexed to an agent A and a time t, Wedgwood sketches a conceptual role semantics as follows:

The semantic value of the practical 'ought'-operator ' $O_{\langle A, t \rangle}$ ' is the weakest property of a proposition p that makes it the case that A is correct to make p part of her ideal plan about what to do at t, and incorrect for A to make the negation of p part of her ideal plan. (p.100)

So when an agent forms the judgment that she ought to F at t, she will be correct in incorporating her F-ing at t into her ideal plan, and incorrect in incorporating her not-F-ing into her ideal plan. (An *ideal* plan is a plan that is not constrained by ignorance or uncertainty (p.97); we shall see the need for this shortly.) What it is to correctly make something a part of one's ideal plan is itself cashed out in terms of a correct plan:

It is correct to make a proposition p part of your ideal plans about what to do if and only if p is logically entailed by a fully correct plan about what to do at p. (p.102)

(A *fully* correct plan is again one that is not constrained by ignorance or uncertainty (pp. 101–2); which seems to be just the same as an ideal plan.)

It will aid exposition if I put these two ideas together to arrive at what I shall call the 'ought-plans' biconditional, since it links, on the one hand, what an agent ought to do, with, on the other, the agent's ideal plans:

An agent ought to Φ at t iff she would be correct in incorporating her Φ -ing at t into her ideal plan, and incorrect in incorporating her not- Φ -ing at t into her ideal plan.

The idea expressed by the ought-plans biconditional is at the heart of Wedgwood's proposal for providing an account that captures the requirement of internalism. I doubt though that, understood as I have understood it, it is strong enough to capture the internalism constraint.

Consider first the conceptual role semantics sketched above for conjunction. The introduction rule for conjunction is couched in terms of permissibility, and rightly so. One is *allowed* to form 'P & Q' from 'P' and 'Q', but one is not *required* to do so. If there were a requirement, that would lead to an immediate regress, since having arrived at 'P' and 'Q' and 'P & Q' one would be obliged to go on to infer 'P & (P& Q)' and 'Q & (P& Q)' and so on.

Now consider the analysis of 'ought' given above. That too seems to be phrased in terms of permission: one would be correct to make the contents of an ought judgment part of one's

plan, incorrect to make its negation part of one's plan. That gives conditions on what is and isn't permitted. There is no requirement that one should incorporate the content of the ought judgment into one's plan. But in this case, the account is too weak. For Normative Judgment Internalism does articulate a requirement. Once one judges that one ought to do something, it requires that one does form the intention to do it. So the semantics for 'ought' doesn't appear to deliver what is needed to secure internalism.

To make the point more vivid, consider Hamlet. This is not quite Shakespeare's Hamlet, but a simpler, practically indecisive, soul. Once this Hamlet is convinced that he ought to do something, he never forms a plan to do its negation. But he frequently fails to form any plan at all. So often the opportunity to do what he judges he ought to do simply passes him by as he drifts, planless. Hamlet does not appear to meet the internalism requirement. Yet he does nothing to violate the semantic condition that Wedgwood places on 'ought' judgments, since he never plans to do otherwise.

Is Hamlet possible? It might be held that, *de facto*, he forms plans contrary to his 'ought' judgments. Even if all that he plans to do is nothing, isn't that still a plan? I think not. There are ways of failing to act that do not involve planning not to act. Sleeping through is one. Less dramatically, failing to think is another. If much current research in social psychology is right, very many of our actions are performed automatically, without the need for planning.

The point remains even when we acknowledge that, on Wedgwood's account, a plan falls far short of an intention (p. 105). Plans include what is entailed by an intention, so that I can plan on sleeping better (say, by intending to take more exercise), even if a straightforward intention to sleep better would be self-defeating (I would lie awake trying to implement it). But Hamlet doesn't form plans that entail doing other than what he judges best. He forms no plans at all.

How could we avoid the problem that Hamlet poses? The obvious approach is to amend the semantic account so that it does not merely grant *permission*, but rather places a *requirement*. Thus we might say:

The semantic value of the practical 'ought'-operator ' $O_{\langle A, t \rangle}$ ' is the weakest property of a proposition p that makes it the case that A is required to make p part of her ideal plan about what to do at t , and required to not make the negation of p part of her ideal plan.

And from this we get a revised version of the ought-plans biconditional

An agent ought to Φ at t iff she is required to incorporate her Φ -ing at t into her ideal plan, and required not to incorporate her not- Φ -ing at t into her ideal plan.

This does avoid the problem, but at an obvious cost. For now we are explicitly talking in terms of requirements, that is of obligations, of things that ought to be done. And that takes us round in a fairly tight circle. Is this a circle that Wedgwood needs to avoid? I think not. Certainly we don't get anything like a reductive definition of 'ought', but then Wedgwood was never after that. But it is not vacuous. Indeed, it is substantial enough that some real objections can be raised. Before that though, a digression.

DIGRESSION: HAVE I UNDERSTOOD WEDGWOOD PROPERLY?

I think that what I have just given is the natural way to understand Wedgwood; certainly it is the way that I understood him initially. But there is another way of reading him according to which he is already endorsing something like the requirement account of 'ought' that I have been arguing he should endorse. In the interest of fairness, and since I think that the ideas in play have a great deal of interest in their own right, let me spell this out.

The point turns around the notion of correctness. As we have seen, correctly making something a part of one's ideal plan is cashed out in terms of a fully correct plan:

It is correct to make a proposition p part of your ideal plans about what to do if and only if p is logically entailed by a fully correct plan about what to do at t . (102)

Now the right hand side of this biconditional is an existential. And the reading that Wedgwood seems to want is that if there is *any* ideal plan that logically entails that p , then you are correct to make p part of your ideal plan, and conversely if there is no ideal plan that logically entails p , then it is incorrect to make p part of your ideal plan. As evidence that that is what he wants note that he now formulates something like the ought-plans biconditional as follows:

For any proposition p , ' $O_{\langle A, t \rangle}(p)$ ' is true just in case there are ideal plans for A to have about what do to at t that logically entail p , and no such ideal plan that logically entails the negation of p (p.102, q.v. p.108)

Since there is an ideal plan in which Hamlet does what he ought to do, and none in which he doesn't, he is correct in forming the plan to do it, and incorrect in failing to do so. So on this understanding, the ought-plans biconditional, appearances notwithstanding, does not merely permit Hamlet to form a plan; it requires him to do so.

But that shows that the initial definition of correctness is rather misleading. If one is taking an exam, there are three things that one can do: give a correct answer, give an incorrect answer, or fail to give an answer at all (some marking schemes penalize the second more than the third to discourage guessing). Similarly it seems that plans should allow three possibilities: correctly incorporate p ; incorrectly incorporate p ; fail to incorporate p at all, and hence be neither correct nor incorrect with respect to p . But on this second understanding of Wedgwood's account of correctness, there is no space for this third possibility. And that's because he's not talking about whether or not something that you have incorporated is correct; he is talking about the correctness or not of incorporating something.

There is an ambiguity here depending on something like the scope of the correctness. The sentence:

It is correct to make a proposition p part of your ideal plans about what to do if and only if p is logically entailed by a fully correct plan about what to do at t (p.102)

could mean either:

- (i) if you make p part of your ideal plans, what you have done is correct iff p is logically entailed by an ideal plan; or
- (ii) the correct thing to do is: make p part of your ideal plan iff p is logically entailed by an ideal plan.

Suppose p is entailed by an ideal plan, but you do not make it part of your ideal plan. Then you have violated the second reading, but not the first. The first one gives you permission to make p part of *your* ideal plan if it is part of *an* ideal plan; the second states that you ought to do so.

None of this affects the main point, which is that Wedgwood needs to understand the semantic account as providing a requirement, rather than permission; the issue has simply been whether he already does, that is, whether the two versions of the ought-plans biconditional are equivalent. End of digression

FROM THE SEMANTICS TO A LOGIC

Having linked plans to ‘ought’ statements by the ought-plans biconditional, Wedgwood uses the constraints to which ideal plans are subject (most centrally a requirement that plans be realizable) to argue that the correct logic for ‘ought’ is standard deontic logic (SDL), that is, the deontic version of the modal logic $\kappa\mathcal{D}$, with the box understood as an obligation operator (pp. 109–116). This is a bold move, since standard deontic logic has come under a great deal of criticism. Indeed both of the distinctive axioms, κ and \mathcal{D} , have been widely held to give rise to irresolvable problems. The κ axiom (symbolizing ‘It ought to be that’ by ‘O’)

$$\text{K: } O(p \rightarrow q) \rightarrow (Op \rightarrow Oq)$$

has given rise to problems such as Chisholm's paradox, and the Paradox of the Gentle Murder, which result from the attempt to capture the further obligations that arise when it is given that the agent will not do the optimally good thing (what ought I do *given* that I am going to fail in one of my obligations?). And the \mathcal{D} axiom

$$\text{D: } Op \rightarrow \neg O\neg p$$

rules out, as a matter of logic, the possibility of contradictory obligations, and hence the possibility of logical dilemmas that could give rise to them.¹ So it is certainly striking if considerations about the logic of plans, together with the ought-plans biconditional, are powerful enough to show these worries to be misplaced.

Actually this is somewhat misleading, since, despite his avowed defense of standard deontic logic, Wedgwood actually adopts a rather different logic, in two respects. First, as we have seen, Wedgwood does not use a simple obligation operator as in SDL, but an operator indexed to an agent and a time; and as he points out, this appears to give him a way of avoiding some of the problems raised for it.

Second, with rather less fanfare, Wedgwood argues in a footnote that the best way to handle the problems stemming from the κ -axiom such as Chisholm's paradox is to add to the logic a primitive deontic conditional along the lines of that proposed by Feldman (p.116 n. 9).² If Wedgwood really does want to follow Feldman all the way here, then this involves a fairly major revision to SDL, for the conditional that is added is a specifically deontic one, the

¹ For an excellent review of the problems surrounding SDL see Paul McNamara's entry ‘Deontic Logic’, and the various supplements to it, in the *Stanford Encyclopedia of Philosophy*.

² Fred Feldman, *Doing the Best We Can* (Dordrecht: Reidel, 1986), and ‘A Simpler Solution to the Paradoxes of Deontic Logic’ *Philosophical Perspectives* 4 (1990) 309–41.

truth conditions of which are determined by a new ordering over the deontic acceptability of worlds.

The resulting logic is thus some way from SDL; indeed, as Wedgwood notes, in its combination of an operator indexed to time and agent, and a primitive deontic conditional it resembles very closely the overall account proposed by Feldman. Obviously I cannot attempt to provide a wholesale evaluation of such a logic here; but we can look at little more closely at the question of how much it can be defended by considerations stemming from the logic of plans.

Let us start with an example discussed by Wedgwood: a case in which there are three options open to an agent, one of which will be really good, one really terrible, and a third merely passably good, though not as good as the first (pp. 30–1). For instance, to take an example given by Frank Jackson, a doctor might have access to three treatments, one of which would cure their patient, one of which would kill them, and one of which would greatly improve their condition without effecting a complete cure.³ Clearly it seems that the doctor ought to give them the first treatment, and equally clearly it seems that doctor ought to form the plan to do so. But now let us add the complication that the doctor does not know which is the treatment that will cure and which is the treatment that will kill; the only treatment they can identify is the treatment that will ameliorate. Now whilst there is a clear sense (an ‘objective’ sense) according to which the doctor ought to give the curing treatment, they would surely be wrong, given the risk of killing the patient, to plan to give it. Given the risks, the thing the doctor ought to plan to do is to give the third, ameliorating treatment.

Now of course one might insist that there is also a sense, a subjective sense, in which the doctor ought to give the ameliorating treatment. But this will not help Wedgwood’s account, since Wedgwood is explicit that it is the objective sense of obligation that he is after: what one ought to do absolutely, not what one ought to do relative to one’s knowledge. And in that sense the case provides a counter-example to the ought-plans biconditional in each direction. For we have something that the agent (objectively) ought to do (give the curing treatment) that they ought not plan to do; and we have something that they ought to plan to do (give the ameliorating treatment) that they (objectively) ought not do.

Wedgwood’s response to this is to restrict the account to propositions about which the agent is fully informed. It is not enough, of course, that they make no mistake; that would only ensure that they made no bad plans. If they are to make the correct plans whenever they have an obligation, they will also need to avoid ignorance. That is, the account will apply only to agents who are relevantly omniscient. (Hence the need to talk of fully correct plans in the discussion above.) Now whenever an account is restricted to idealized agents there is a question of how relevant it is to actual agents. But there is a particular question that arises in the context of a planning theory. For there is good reason to think that much of the point of forming plans is to enable agents to overcome their cognitive limitations. Once we abstract to omniscient agents, there is a real question about whether there is a need for plans at all.

Why do ordinary mortals need to form plans? Why shouldn’t they simply act on their judgments of what is best at the time? Much of it comes down to the need to coordinate in the face of ignorance. Since my cognitive resources are limited, I cannot keep deliberating until I am certain that I have arrived at a judgment about what is best. I need to decide what

³ Frank Jackson ‘Decision-Theoretic Consequentialism and the Nearest and Dearest Objection’ *Ethics* 101 (1991) 461–82

to do on the basis of reasonable consideration; and then, since I and others will need to coordinate around that decision, I will need to keep it relatively stable.⁴

If that is the basis of our need for plans as we make them, then the move to relevantly omniscient agents is the move to creatures who have no need of them. Perhaps they will have plans that perform some other function; but if so, it is quite unclear that the norms that will govern them will be the norms that will govern plans as we have them. At the limit they will simply be the norms that govern what they ought to do; but in that case, the link provided by the ought-plans biconditional will be uninformative, since the plans will be the shadows of the oughts. The conception of planning will not introduce any independent constraints, since we will not be dealing with an independent conception of planning.

I think that Wedgwood comes perilously close to such a position. Consider his comments on another case that has been widely held to show a gap between what one ought to do and what one ought to plan to do: Kavka's Toxin Case.⁵ You are offered a large sum of money to form the intention to drink a toxin that is not dangerous but will make you very sick for a short time. You calculate that the sum is so great that it would be worth enduring the sickness to gain it. But you are not required to actually drink the toxin: you will receive the money simply for forming the intention to do so, an intention which you can abandon before the time for drinking comes. Much of the debate surrounding the case has focused on whether a rational agent would be able to form the intention and win the money. We needn't pursue that though. We need only draw the plausible conclusion that the case provides a reason for forming the intention, to drink the toxin, whilst at the same time not providing a reason for drinking it.⁶ Thus it provides a case in which what one ought to do, and what one ought to plan to do, come apart.

Wedgwood's response to this is to say that there are two senses in which forming the plan to drink the toxin is the correct thing to do: in one sense it is correct to form it; in the other, which is the sense he is after in his account, it is not. Wedgwood draws an analogy with belief: if someone offered you a large sum of money to believe something harmless that you knew to be false, would it be correct to believe it? In one sense, yes, since you would gain the money at little cost. In another no, since the norm governing belief requires that your beliefs be true.

Wedgwood's response here is revealing. He is surely right that the norm governing belief does not allow you to form a belief in such circumstances (though quite how to spell out the norm is controversial). But what is the parallel norm governing the forming of plans? The obvious candidate is that one ought only to plan to do that which one ought to do. However, as the toxin case seems to show, our ordinary conception of a plan isn't governed by such a norm; and insofar as we can come up with a conception of a plan that is, it isn't driven by considerations that stem from a distinctive notion of planning. Rather, it is driven by considerations that stem from the notion of what one ought to do.

Let us return to the issue of the right logic for obligation. Wedgwood's striking idea, recall, was that one could shed light on the logic of obligation by looking at the constraints

⁴ See Michael Bratman *Intention, Plans and Practical Reason* (Cambridge MA: Harvard University Press, 1987) for the origins of this approach. The idea that the need for the will in general results from our imperfections is an ongoing theme of my *Willing, Wanting, Waiting*, (Oxford: Clarendon Press, 2009).

⁵ Gregory Kavka, 'The Toxin Puzzle', *Analysis* 43, (1983) pp. 33–6.

⁶ I'm actually rather skeptical of this conclusion: I suspect that the case provides reason to drink the toxin, but I shan't indulge my unorthodox views here. See *Willing, Wanting, Waiting*, Ch 7.

on planning. My worry has been that, by making the idealizations he has made, Wedgwood is not really introducing constraints stemming from the notion of planning. Rather the considerations that he adduces are those that were already there in the idea of obligation. In closing I want to suggest that when we really do bring in independent considerations from the notion of planning, we should be reluctant to import them into our idea of obligation; that is, when the ought-plans biconditional is substantial, we have reason to reject it.

To see this, consider what Wedgwood's method might have to say about the second controversial axiom of standard deontic logic, the D-axiom, ensuring consistency of obligation. Here at least we might hope that considerations from the theory of planning would provide support. For if we think that our plans should be consistent, won't this transfer through the ought-plans biconditional to give a consistency constraint on 'ought'? Even here I am unconvinced. In so far as we are tempted by the idea of dilemmas that give rise to conflicting obligations (whether moral or practical), why should we abandon this just because we recognize that these cannot be fulfilled? It is surely right that if one has conflicting obligations, it would typically be irrational to try to fulfill both: as a practical matter that would typically lead to disaster.⁷ But that is a distinctive feature of the logic of planning. It is a complacent doctrine to think that that in itself guarantees that there can be no dilemmas. Of course there is a familiar position that says that there can be no obligation to do what is impossible. But that is a consideration from the side of the theory of obligation, and it is exactly that doctrine that the person who believes in dilemmas will reject.⁸

My comments here have been largely critical; that is the nature of a critical response. It should go without saying, though I will say it, how much I have benefitted from wrestling with Wedgwood's intriguing proposals. I am sure it will be highly influential, as it certainly deserves to be.⁹

⁷ I personally think that there are times when it can be rational to form contradictory plans: times when one is uncertain which will be successful (see *Willing, Wanting, Waiting* Ch 2). This is not immediately relevant to the issues here, so again I shall not pursue it. But it does, I think, give further evidence that the norms governing intentions have a distinctive practical cast.

⁸ Indeed, my suspicion is that in accepting the logic of obligation that he does, Wedgwood has already moved to thinking that the norms that govern obligation are quite distinct from the norms governing practical deliberation: for I cannot see that the deontic conditional that he introduces to avoid problems like Chisholm's paradox has a distinct practical analogue. But issues here are complicated.

⁹ A preliminary version of this piece was presented at the symposium at the Western Division of the APA in spring 2009, with Ralph Wedgwood and Peter Railton. I learned a lot, but doubtless not enough, from Wedgwood's response then.