The Onus in 'Ought'

MAHRAD ALMOTAHARI & BRIAN RABERN

- forthcoming in *Analysis* -

1. Obligations are requirements and requirements are necessities. Naturally, then, obligations are necessities. This compelling syllogism provides an important clue to the logic of obligation: it's simply the logic of necessity, suitably restricted.

Fortunately, the logic of necessity is well understood. If we interpret the standard notation of modal logic in the appropriate way, letting box abbreviate 'It ought to be that' and diamond 'It is permissible that', then we can fiddle around with our Kripke models to formulate a complementary semantic theory (Montague 1960). Incorporating familiar tools for the representation of context sensitivity—specifically, a modal base and an ordering source—enables the logic of necessity to serve as a semantic theory for a rich and philosophically interesting fragment of natural language (Kratzer 2012). The precision and uniformity of this approach, coupled with its empirical successes over several decades of fruitful research, offers a powerful reason to think that little more is required. Perhaps we should expect some fine-tuning here and there, but future revisions will preserve the foundational assumption that the grammar and meaning of 'ought' is a special case of the grammar and meaning of box. Portner (2009) and Chrisman (2016) call this the standard theory of deontic modals.

To avoid equivocation, boxes and diamonds are often reserved for the expression of "alethic" modalities. Deontic necessity is usually expressed with an 'O' (for 'ought' or 'obligation'). Consequently, the structure implicit in (1) is made explicit in (1*).

- (1) Nobody ought to die of hunger.
- (1*) O(Nobody dies of hunger)

A sentence of the form 'O φ ' is true at a context iff the relevant domain of possible worlds in that context (roughly speaking, the modal base) is such that, in all the highest ranked worlds relative to the operative standards in that context (aka the ordering source), φ is true.

The simplicity of this analysis is attractive. But its scope is controversial. There's an undercurrent of dissatisfaction stemming from the sense that 'ought', as a propositional operator, is most apt for expressing how things ought to be, not what someone ought to do.¹ Numerous challenges suggest, in various ways, that the standard theory misses something important about how obligation is related to a particular agent.² Some 'ought'-statements don't just say how things ought to be, they also seem to indicate where the onus to act lies. Can the standard theory do justice to the representation of onus in 'ought'-statements? –No, it can't. And this raises a puzzle about the direction of future theorizing.

2. Our central question is suggestive, but perhaps not sufficiently clear. We'll try to bring it into sharper focus with an example.

¹ See, e.g., Humberstone (1971; 1991), Harman (1973), Williams (1981, ch. 9), Geach (1982), Schroeder (2011), Broome (2013, ch. 2).

² One challenge is based on observations about iterated deontic modality (Barcan Marcus 1966); another on passivization (Humberstone 1971, 1991; Harman 1973; Geach 1982; Schroeder 2011); a third on conditional obligations and plural agency (Ross 2010); and perhaps the most recent challenge relies on verb raising and control (Schroeder 2011). The challenge from passivization is highly relevant for our discussion. Going forward, we'll focus on it. But one referee asks how the central argument of Section 3 differs from the challenge in Ross (2010). Briefly, our argument doesn't require agglomeration; Ross's does. Furthermore, Ross acknowledges that his challenge might be met by analyzing *ought-to-dos* in terms of conditional *ought-to-bes* (pp. 316-317); our argument isn't even potentially vitiated by this strategy. Spot the standard theorist her favorite conditional *ought-to-be* analysis and our problem remains.

- (2) John ought to kiss Mary.
- (2*) O(John kisses Mary)

Sentence (2) attributes an obligation to John and thereby establishes a basis for holding him accountable. Equivalently, it places the onus on him to kiss Mary. The most salient reading of (3), however, is neutral with respect to where the onus lies.

- (3) Mary ought to be kissed by John.
- (3*) O(Mary is kissed by John)

Perhaps John has promised his wife that he won't kiss Mary, in which case he would have an obligation to refrain from doing so. Still, you might think that Mary deserves to be kissed by John. After all, she won the competition, and the prize, we were all told, would be a kiss from John. It would be unjust if Mary isn't given what she's owed. So, it ought to be that Mary is kissed by John.

The difference between (2) and (3) indicates that they express different contents. But their respective analyses, according to the standard theory, appear to be equivalent, since (i) the embedded sentences appear to be equivalent and (ii) the univocality of 'O' is baked into the standard theory.

Some have abandoned the view for this reason. There are 'ought'-claims that might be analysed in the standard way, as a propositional operator combined with a (declarative) sentence. But, for onus-attributing 'ought'-statements, new technology is required. A simple idea would be to index the propositional operator to different agents: ' $O_a\varphi$ ' versus ' $O_b\varphi$ '. Or one might rely on a two-place relation between an agent and a proposition: ' $O(a,\varphi)$ ' versus ' $O(b,\varphi)$ '. Alternatively, one might opt for a relation between an agent and an act-type, or an agent and a property, or an agent and a centered proposition. These differences don't matter for our purposes. What matters is that these are all ways of moving away from the standard theory. For now, our interest lies in whether there's good reason to undertake this movement, not in the specific direction it might lead.

The contrast between (2) and (3) doesn't, by itself, undermine the standard theory. One could question the assumption that the embedded sentences are equivalent, or that they're equivalent on their intended readings. One very natural reading of 'John kissed Mary' is agential; that is, kissing Mary is something that John is said to have done; it's not something that's merely said to have happened to him. A natural reading of 'Mary was kissed by John', however, is nonagential; kissing Mary is something that's merely said to involve John by accident. If the kissing is an action that John performs at all, then it was an unintentional action. Given this intuitive contrast, the standard theorist might insist that the embedded sentences in (2) and (3) are, at most, equivalent on a disambiguation. One standard framework in which this contrast can be represented is stit theory (Belnap and Perloff 1988). The acronym, 'stit', stands for the phrase sees to it that, and one of the central ideas motivating the framework is that the difference between agential and non-agential readings is broadly a matter of logic. So there must be a formal difference between the embedded sentences about John and Mary—one that we can represent with the phrase 'x sees to it that', or 'stitx'.

According to stit theory, on the relevant disambiguations, the embedded sentence in (2) instantiates the form given in (2^{**}) , and the embedded sentence in (3) exemplifies the structure in (3^{**}) .

The formal difference between (2**) and (3**) is supposed to reflect the semantic distinction between agential readings of NP-VP constructions and their non-agential counterparts. Note that (2**) implies (3**) but not vice versa. The difference between (2**) and (3**) is, basically, the difference between intentional and unintentional action.

Other frameworks explain the semantic difference between (2) and (3) differently. According to the contrastivist approach of Finlay and Snedegar (2014), (2) and (3) differ because they contextually invoke different contrast classes. (2) is true just in case it ought to be that John kisses *Mary* rather than that John kisses *Alice* or that John kisses *Bernadette*. Because the agent, John, remains constant in each case, (2) is agential. (3), however, is true just in case it ought to be that *John* kisses Mary rather than that *Calvin* kisses Mary or that *Dwight* kisses Mary. Because the agent varies, (3) is non-agential.⁴

Our summary of stit theory and contrastivism indicates how resourceful the standard theory is. So you might think the answer to our central question—can the standard theory do justice to the representation of onus in 'ought'-statements—is yes.⁵ But there's a more powerful challenge in the offing. Our aim is to present it and explain why it matters.

- 3. The challenge begins with an influential principle from ancient Greek ethics: it's better to be wronged by another person than to wrong them. Alternatively:
 - (4) One ought to be wronged by another person rather than wrong them.

The basic idea is that wronging someone is ultimately damaging to oneself, and the damage to one's character/wellbeing is so costly that one ought to endure a wronging rather than be its agent.⁶ Perhaps that's why Christ recommended that his followers turn the other cheek, and why Gandhi practiced *satyagraha*. The idea seems to resonate across various times and cultures. And a familiar conception of the relationship between character and wellbeing seems to lie behind it, a conception that students encounter in Plato's *Republic*, according to which the soul of an unjust person is in turmoil. In other words, immorality is deeply incompatible with human flourishing.

Arguing for (4) would be a monumental undertaking, and not one that we're able to attempt. Fortunately, our argument doesn't require that (4) be true; it only requires that it be deontically coherent. And we trust that its role in the history of philosophy demonstrates that much. More precisely, we assume that nothing of the form ' $O\varphi \wedge O\sim\varphi$ ' is a consequence of (4). For what it's worth, when we discuss (4) with friends and colleagues, no one expresses a worry to the effect that it imposes unsatisfiable demands.

⁴ The full story involves an interaction effect resulting from (i) contextually supplied contrast classes, (ii) a constraint on the kinds of contrast classes that permit agential readings, and (iii) apparent violations of a Gricean conversational maxim: *Avoid prolixity!*

³ Schroeder (2011) criticizes this response to the passivization challenge. But Chrisman (2012) and Finlay and Snedegar (2014) argue that his criticism is based on a mistaken conception of the linguistic data. An additional challenge to Schroeder, from a somewhat different direction, appears in Lee (2021).

⁵ Broome (2013, ch. 2) criticizes the use of 'stit' to represent onus. His criticism relies on an intuition about a hypothetical case (pp. 20-22). Our intuitions about the case are equivocal. We're happy to go whichever way the most promising overall theory dictates.

⁶ (4)'s role in ancient Greek thought is discussed in Müller (1977), an undeservedly neglected paper from which we borrow the example. Müller credits Democritus for the principle and relies on it to argue against utilitarianism. We don't think his argument works—certainly not as he states it. Utilitarianism doesn't require the standard theory of deontic modals. A utilitarian might acknowledge that 'ought' has a use on which it's a predicate modifier and not a sentential operator. It's the standard theory for which the ancient Greek principle makes trouble, as we'll show momentarily. To be fair, the standard theory wasn't in view when Müller wrote his paper, though he does show an awareness of relevant trends in deontic logic (p. 118, fn. 17).

Now, assume for simplicity that 'rather than' is a binary sentence connective: ' $\varphi > \psi$ '. How, then, does it interact with 'ought'? An example will help.

- (5) Sam ought to finish his homework rather than go out all night.
- (5*) O(Sam finishes his homework > Sam goes out all night)

According to the standard theory, (5) is true only if, among the worlds where Sam either finishes his homework or goes out all night, Sam finishes his homework in the highest-ranked worlds. More generally, 'O($\varphi > \psi$)' is true only if, among the φ -or- ψ -worlds in the relevant domain, the highest ranked worlds are φ -worlds and not ψ -worlds.⁸ So, the standard theory yields the following analysis of (4):

$$(4^*) \ \forall x \forall y [x \neq y \supset \mathcal{O}(\mathbf{W}yx > \mathbf{W}xy)]^9$$

But we claim that (4^*) is incoherent. This becomes apparent when we consider two of its immediate consequences. (We assume that a and b are distinct agents.)

- (6) O(Wba > Wab)
- (7) O(Wab > Wba)

According to the standard theory, (6) is true only if, among the Wab-or-Wba-worlds, the highest ranked ones are Wba-worlds and not Wab-worlds. But likewise, (7) is true only if, among the very same worlds, the highest ranked ones are Wab-worlds and not Wba-worlds. Thus, (4*) is true only

$$(4^{**}) \forall x \forall y [x \neq y \supset OWyx > OWxy]$$

Two points are relevant. First, assuming a textbook model of ellipsis—one on which it applies only to syntactic constituents that mirror maximal projections—there's evidence against the elided-'ought' hypothesis. Consider:

- (i) A student ought to study hard for her exam rather than cheat on her exam.
- (1*) A student₁ ought to study hard for her₁ exam rather than a student₂ ought to cheat on her₂ exam.

If (i) were derived from (i*) in the way that (4) might be alleged to derive from (4**), then (i) ought to have a reading on which it's equivalent to (i*). But there's no such reading. (For relevant discussion, consult Sailor and Thoms (2014).) Additionally, even if (4**) is a viable LF, it doesn't shield the standard theory from our criticism. The reason is that the elided 'ought' would require the same interpretation as its antecedent (Wilder 1997, p. 72ff.). For example: both (ii) and (iii) could be uttered truthfully, but there's no way to make the semantic content of (iv) true—even allowing intrasentential context shifts.

- (ii) (In view of the laws governing the US during the early 60s) Muhammad Ali ought to fight in Vietnam.
- (iii) (In view of the Shariah) Muhammad Ali ought to refrain from fighting in Vietnam.
- (iv) Muhammad Ali ought to fight in Vietnam and refrain from fighting in Vietnam.

(We gratefully acknowledge help from Craig Sailor and two anonymous referees.)

⁷ See Thompson (1972) for discussion of 'rather than'-clauses. It would be natural to treat '>' as material nonimplication: ' $\varphi > \psi$ ' is true iff $\varphi \wedge \sim \psi$, but with a presupposition to the effect that $\varphi \vee \psi$. Although this analysis combines with the standard theory of 'ought' in a way that generates the reading of 'O($\varphi > \psi$)' we rely on above, it raises questions that require further investigation. In any case, we aren't committed to a particular theory of 'rather than'.

⁸ Finlay and Snedegar (2014) and Snedegar (2015) have this analysis in mind. Given the limit assumption, and a connected ordering relation, this analysis is equivalent to Kratzer's interpretation of ' φ is a better possibility than ψ ' (Kratzer 2012, p. 41). Our argument can be reformulated accordingly. (We thank Kai von Fintel and an anonymous referee for discussion.)

⁹ Could (4) be derived from an LF containing a second 'ought'?

if some world is ranked higher than itself. It is, therefore, contradictory. But (4) is coherent. It certainly *feels* like wisdom. But we won't press the point. All that matters is that a principle as influential to the history of ethics as (4) shouldn't be made incoherent by developments in formal semantics.

Note that (4) is explicitly contrastive in form, and the problem issues from the logical consequences of (4*). These elements of the argument bar a potential resource for the standard theorist to exploit (à la Finlay and Snedegar), namely, contextually supplied contrast classes and Gricean pragmatics. Furthermore, sprinkling in 'stit' operators to highlight the locus of the agential reading also doesn't help capture the intended interpretation. 11

Aficionados will probably notice that our central argument subtly relies on agglomeration: $O\varphi$; $O\psi \models O(\varphi \land \psi)$. The standard theory validates this principle (Kratzer 2012), but it's sometimes seen as a shortcoming (Cariani forthcoming). There are ways of rejecting agglomeration while preserving the foundational architecture of the standard theory (Swanson 2011). So, you might ask, doesn't our argument have a limited bearing on the theory's viability? –No. Even if we abandon agglomeration, (6) and (7) will entail a deontic conflict. This is because ' $O(\varphi > \psi)$ ' implies ' $O\sim(\psi > \varphi)$ '. (If Sam ought to finish his homework rather than go out all night, then he oughtn't go out all night rather than finish his homework.) So (4*) yields the following: $O(Wab > Wba) \land O\sim(Wab > Wba)$. But, as we observed at the outset, (4) doesn't commit one to a deontic conflict. To be clear, we're not now claiming that deontic conflicts are impossible; only that (4) remains silent about them.¹²

There may be some lingering suspicion that our argument trades on a peculiarity of 'rather than'-constructions or generic 'one', and that a more adequate theoretical understanding of these constructions will resolve the difficulty, leaving the standard theory intact. Though these expressions raise interesting issues that deserve attention, we think they're dialectically irrelevant here.

(8) If a person will either be wronged by another or wrong them, then they ought to be wronged by the other and not wrong them.

Sentence (8) expresses basically the same moral principle, but without the use of 'rather than' or 'one'. The indefinite, 'a person', has universal force, akin to 'a farmer' in more familiar donkey sentences: "If a farmer owns a donkey, he beats it" (Heim 1982). Given (8), the argument against the standard theory would proceed as before, yielding the same conclusion.

Our challenge, then, is this: how can the standard theorist accommodate the coherence of both (4) and (8) while maintaining the univocality of 'ought' as an unrelativized operator on declarative sentences? We do not say it can't be done, but we do say it isn't obvious how.

_

Than $(\varphi, \psi) \wedge$ Better-Than (ψ, φ) ' is also clearly unacceptable.

¹⁰ That (4) should pose a problem for contrastivism is particularly interesting, since it's sometimes taken for granted that "contrastivism easily handles claims which include an explicit 'rather than' clause" (Snedegar 2015, p. 4).

¹¹ For example, consider ' $\forall x \forall y [x \neq y \supset O(\text{stit}_y Wyx > \text{stit}_x Wxy)]$ ', which has these immediate consequences:

^(6*) $O(\operatorname{stit}_b Wba > \operatorname{stit}_a Wab)$

^(7*) $O(\operatorname{stit}_a Wab > \operatorname{stit}_b Wba)$

According to the standard theory, (6*) is true only if the highest ranked worlds (among the relevant worlds) are stit_{b} -Wba-worlds and not stit_{a} -Wab-worlds. But (7*) is true only if (among the same worlds) the highest ranked ones are stit_{a} -Wab-world and not stit_{b} -Wba-worlds. Like (4*), this is contradictory. Other attempts can avoid contradiction but at the cost of misconstruing (4). For example, both ' $\forall x \forall y [x \neq y \supset O(\operatorname{stit}_{x} Wyx > \operatorname{stit}_{x} Wxy)]$ ' and ' $\forall x \forall y [x \neq y \supset O(Wyx > \operatorname{stit}_{x} Wxy)]$ ' are consistent. The former recommends masochism: it's better to see to it that someone wrongs you than to wrong them. The latter misses the fact that (4) concerns both intentional and unintentional wronging.

12 Variations on (4) incorporate deontic 'better than' or deontic 'must'. With the latter, agglomeration is usually assumed; so 'must $\varphi \land$ must $\sim \varphi$ ' requires a contradiction. Given the logic of 'better than', a result such as 'Better-

The problem for the standard theory is that, by its lights, the interpretation of 'O' is insensitive to the source of a wronging; it only cares about how worlds are ranked. So, it shouldn't matter whether it's a wronging b or b wronging a. Either way, there's a wronging. This sort of insensitivity is perfectly fine when one wants to say how circumstances ought to be, but inadequate when one's meaning concerns what one ought to do (where the onus in 'ought' lies). The use of 'stit' and similar devices may help somewhat with this issue, as our brief discussion of passivization suggests, but (4) and (8) are cases where these devices don't help.

This diagnosis presents us with a choice. One option is that 'ought' is ambiguous. On one reading it semantically expresses a relation between agents and actions (or centered worlds or...); on the other, a function from propositions to truth-values. This bifurcation of meaning might be formally implemented in various ways. But if ambiguity is the way to go, then (4)/(8) provide linguistic evidence for its acknowledgment. It certainly wasn't clear at the outset that some form of semantic bifurcation would be necessary. As Chrisman (2016) observes, "The only attempt I know of by philosophers sympathetic to [this] kind of bifurcation ... to provide *linguistic* argument for an agentive versus nonagentive distinction in the semantic functioning of the world 'ought' is Schroeder's (2011)" (p. 127, emphasis in original). If, in the end, one opts for ambiguity, the argument that we've offered in this paper vindicates the decision on empirical linguistic grounds.¹³

But bifurcation isn't cost free. If the same pattern is seen to occur cross-linguistically, and with other modals (e.g., with 'may' and 'can'), then positing a brute, one-off ambiguity wouldn't carve meaning at the joints. The challenge that we've raised here motivates the conclusion that—insofar as we maintain the univocality of 'ought'—the semantics of 'ought'-statements must be centered around personal obligation, instead of impersonal rankings of circumstances. In a slogan, 'ought' implies onus. Like the ambiguity strategy, this thought can be formally implemented in various ways (Chrisman 2016; Hacquard 2010). But, as before, we won't take a stand on which implementation is the most promising. We'll simply note a lingering question for this approach: if one's account begins with personal obligation, how will one understand the ought-to-bes that motivated the standard theory? More specifically, how will one analyze the 'ought'-statements that don't seem to place an onus on anyone in particular?

University of Edinburgh 40 George Square, EH8 9JX, United Kingdom

> malmotah@ed.ac.uk brian.rabern@ed.ac.uk

⁻

¹³ This is one respect in which our argument improves on previous attempts to demonstrate onus in 'ought'. For instance, Broome (2013) is concerned not with empirical linguistics, but conceptual regimentation—a largely stipulative enterprise that's meant to advance normative inquiry, not formal semantics. Furthermore, our argument makes do with much less: we don't assume that deontic conflicts are impossible; we don't assume a controversial closure principle; and we don't assume any theory of propositions.

¹⁴ As far as we can tell, our puzzle puts considerable pressure on the position staked out in Lee (2021), according to which 'ought' is univocal, but not a control verb, and therefore not relativized to an agent.

¹⁵ We received valuable help from Matthew Chrisman, Kai von Fintel, Dan Lassiter, Craig Sailor, Wolfgang Schwarz, Will Small, Justin Snedegar, Wataru Uegaki, and two anonymous referees. Special thanks to Anselm Müller for his inspirational work.

References

- Barcan Marcus, Ruth. 1966. 'Iterated Deontic Modalities'. Mind 75: 580-582.
- Belnap, Nuel and Perloff, Michael. 1990. 'Seeing to it that: A canonical form for agentives'. In H.E. Kyburg Jr., R.P. Loui, and G.N. Carlson, eds., *Knowledge Representation and Defeasible Reasoning*. Springer.
- Broome, John. 2013. Rationality through Reasoning. Wiley Blackwell.
- Cariani, Fabrizio. Forthcoming. 'Deontic Logic and Natural Language'. In D. Gabbay et al., eds., Handbook of Deontic Logic Vol. II. College Publications.
- Chrisman, Matthew. 2012. "Ought" and Control'. Australasian Journal of Philosophy 90: 433-451.
- Chrisman, Matthew. 2016. The Meaning of 'Ought': Beyond Descriptivism and Expressivism in Metaethics. Oxford University Press.
- Finlay, Stephen and Snedegar, Justin. 2014. 'One Ought Too Many'. *Philosophy and Phenomenological Research* 89: 102-124.
- Geach, Peter. 1982. 'Whatever Happened to Deontic Logic?' Philosophia 11: 1-12.
- Hacquard, Valentine. 2010. 'On the Relativity of Modal Auxiliaries'. *Natural Language Semantics* 18: 79-114.
- Harman, Gilbert. 1973. Review of Robert Wertheimer, *The Significance of Sense: Meaning, Modality, and Morality* (Cornell University Press). *Philosophical Review* 82: 235-239.
- Heim, Irene. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. Ph.D. Thesis, University of Massachusetts, Amherst.
- Humberstone, Lloyd. 1971. 'Two Sorts of "Ought's'. Analysis 32: 8-11.
- Humberstone, Lloyd. 1991. 'Two kinds of agent-relativity'. Philosophical Quarterly 41: 144-166.
- Kratzer, Angelika. 2012. Modals and Conditionals. Oxford University Press.
- Lee, Junhyo. 2021. "Ought" and Intensionality'. Synthese 199: 4621-4643.
- Montague, Richard 1960. 'Logical necessity, physical necessity, ethics, and quantifiers'. *Inquiry:* An Interdisciplinary Journal of Philosophy 3: 259-269.
- Müller, Anselm. 1977. 'Radical Subjectivity: Morality Versus Utilitarianism'. Ratio 19: 115-132.
- Portner, Paul. 2009. Modality. Oxford University Press.
- Ross, Jacob. 2010. 'Personal and Impersonal Obligation'. Journal of Philosophical Logic 39: 307-323.
- Sailor, Craig and Gary Thoms. 2014. 'On the Non-existence of Non-constituent Coordination and Non-Constituent Ellipsis'. *Proceedings of the 31st West Coast Conference on Formal Linguistics*,

- ed., R. E. Santana-LaBarge. 361-370. Cascadilla Proceedings Project.
- Schroeder, Mark. 2011. 'Ought, Agents, and Actions'. Philosophical Review 120: 1-41.
- Snedegar, Justin. 2015. 'Contrastive Semantics for Deontic Modals'. In M. Blaauw, ed., Contrastivism in Philosophy. Routledge. Page reference to online version.
- Swanson, Eric. 2011. 'On the Treatment of Incomparability in Ordering Semantics and Premise Semantics'. *Journal of Philosophical Logic* 40: 693-713.
- Thompson, Sandra Annear. 1972. 'Instead of and rather than clauses in English'. Journal of Linguistics 8: 237-249.
- Wilder, Chris. 1997. 'Some Properties of Ellipsis in Coordination'. In *Studies on Universal Grammar* and Typological Variation, eds. A. Alexiadou and T. Hall. 59–107. John Benjamins.
- Williams, Bernard (1981) Moral Luck: Philosophical Papers, 1973–1980. Cambridge University Press.