Closing the 'Is'-'Ought' Gap

STEPHEN MAITZEN
Dalhousie University
Halifax, NS
B3H 3J5

I. Introduction

In a dense and fascinating article of some ten years ago, Toomas Karmo adds his voice to the chorus of philosophers who deny the possibility of soundly deriving 'ought' from 'is.'[1] According to Karmo, no derivation containing an ethical conclusion and only non-ethical premises can possibly be sound, where 'sound' describes a deductively valid derivation all of whose premises are true. He also suggests that the only valid derivations of 'ought' from 'is' will be trivial ones. His argument has, to my knowledge, gone unrebutted; on the contrary, it has won recent endorsements, some philosophers evidently believing that he has finally put to rest the issue of the logical autonomy of ethics.[2] Against that trend, I intend to rebut his argument, both by falsifying the taxonomy on which his argument relies and by offering a nontrivial and potentially sound 'is'-'ought' derivation of my own.

---

1 'Some Valid (but no Sound) Arguments Trivially Span the "Is"-"Ought" Gap,' *Mind* 97 (1988): 252-7; subsequent references to this article will give page numbers only. Other members of the chorus include E. J. Borowski, 'Moral Autonomy Fights Back,' *Philosophy* 55 (1980): 95-100; I. L. Humberstone, 'First Steps in Philosophical Taxonomy,' *Canadian Journal of Philosophy* 12 (1982): 467-78, and 'A Study in Philosophical Taxonomy,' *Philosophical Studies* 83 (1996): 121-69; and Charles R. Pigden, 'Logic and the Autonomy of Ethics,' *Australasian Journal of Philosophy* 67 (1989): 127-51; as well as many of the authors cited in Kai Nielsen, 'On Deriving an Ought from an Is: A Retrospective Look,' *Review of Metaphysics* 32 (1979): 487-514; 487 n. 1.

2 For example, Humberstone, 'A Study': 'As Karmo shows, ...no sound arguments with [only] non-ethical premises have an ethical conclusion' (152).

II. Karmo's Taxonomy

Karmo considers the following derivation of an ethical conclusion from two allegedly non-ethical premises:[3]

(A1)   Everything that Alfie says is true.

(A2)   Alfie says that it ought to be the case that everyone is sincere.

Therefore:

(A3)   It ought to be the case that everyone is sincere.

He concedes that A2 is a non-ethical premise and that A3 is an ethical conclusion, but, he says, whether A1 is or is not an ethical premise depends on the contingent matter of whether Alfie has, in fact, asserted any ethical sentences: if Alfie *has*, then A1 is an ethical premise, since anyone who accepts A1 is committed, knowingly or not, to the truth of at least one particular ethical sentence.[4] As Karmo himself puts it, if Alfie has in fact asserted some ethical sentence or other, then A1 is an ethical sentence because its truth or falsity 'turns on a substantive ethical matter' (255).

As we will soon see, Karmo's taxonomy of ethical sentences depends crucially on the notion that some sentences are only *contingently* ethical.[5] One could, of course, reject that notion, but even granting it one can falsify the taxonomy. He develops the notion as follows. Consider the set U containing all and only the uncontroversially ethical sentences, those sentences that would be classified as ethical sentences by all sides in the debate over the logical autonomy of ethics.[6] Although Karmo does

---

3   Page 253. The numbering of the steps is mine.

4   In conformity with Karmo's usage, I will be using the expression 'ethical sentence' rather than what Karmo (252 n. 3) concedes 'might seem more natural,' the expression 'ethical proposition.'

5   According to Karmo, 'the debate [over the logical autonomy of ethics] has suffered from an apparent assumption that if there is a coherent notion of being-ethical, then [being-ethical] is [a noncontingent property] of sentences, rather than [a contingent one]' (252). He cites Humberstone's 'First Steps' as also recognizing that some sentences are only contingently ethical (252 n. 2).

6   He gives these examples. 'It ought to be the case that all New Zealanders are shot' is, he says, uncontroversially ethical and, indeed, 'ethical in every possible world.' 'Everything that Alfie says is true' and the compound sentence 'Either tea drinking is common in England or it ought to be the case that all New Zealanders are shot'

not make the supposition explicit, for the sake of simplicity let us suppose that the membership of U is fixed across all possible worlds (or, more properly, that across all possible worlds U is the set picked out by the description 'the set containing all and only the uncontroversially ethical sentences').

The next step in the development of his taxonomy is the notion of an 'ethical standard,' about which Karmo says this:

> [T]he ethical standard subscribed to by a person is completely determined once it is determined what truth values he assigns to all uncontroversially ethical sentences.... We take it that just as some one possible world is the actual world, so some one ethical standard is the correct ethical standard. When people simply say, "Sentence S is true", we take them to mean "S is true in the actual world with respect to the correct ethical standard". When people simply say, "S is true in world w", we take them to mean "S is true in w with respect to the correct ethical standard" (254).

An ethical standard, then, is any complete — and presumably, although Karmo does not say, consistent — assignment of truth-values to the members of U, the uncontroversially ethical sentences. The final two statements in the quoted passage are peculiar, since they do not specify that S is an *ethical* sentence; instead, they implicate the correct ethical standard in the assertion 'S is true' regardless of S's status as ethical or non-ethical. But let that pass. No doubt one could also raise questions about Karmo's rather thin conception of ethical standards,[7] but his account ends up failing even if we grant him that conception as well.

Now for the relevant taxonomy. Karmo stipulates that 'we define a sentence S to be ethical in a world w just in case S is true in w with respect to one ethical standard, and false in w with respect to another ethical standard' (254). So S is an ethical sentence in a world w if and only if at least one ethical standard (call it 'E') assigns S the value 'true' in w and at least one other ethical standard (call it 'E''') assigns S the value 'false' in w. Karmo evidently does not require that either E or E' be the correct standard (in the actual world or in any other world); a difference in the truth-values assigned to S by E and E' in w is necessary and sufficient for S's being ethical in w '[n]o matter what is, in fact, the correct ethical standard — whether E, or E', or something else altogether' (255). If S

---

are not uncontroversially ethical; they are, he says, ethical in some worlds and non-ethical in others (254-5).

7 Even Karmo's ally Humberstone thinks that 'further discussion of what [Karmo's] "ethical standards" are and precisely how they work is called for' ('A Study,' 156).

should be a noncontingently true, noncontingently ethical sentence,[8] then, for Karmo, the correct ethical standard — suppose it is E — assigns S the value 'true' in every world, and, also in every world, at least one other ethical standard E' assigns S the value 'false'; E' will, of course, be an incorrect standard in every world, but some ethical standard needs to differ from E in its assignment of truth-value to S in a world w in order for S to be an ethical sentence in w.

Again, the basic idea is that the truth of an ethical sentence depends on what the correct ethical standard is, while the truth of a non-ethical sentence does not so depend. If Alfie has asserted some uncontroversially ethical sentence or other, then anyone who assents to 'Everything Alfie says is true' is 'himself taking on an ethical commitment (whether he is aware of this or not)' (255). One may doubt whether Karmo's taxonomy captures the idea of 'taking on an ethical commitment.' Consider, for instance, a sentence that conjoins an uncontroversially ethical clause and an uncontroversially non-ethical falsehood: 'Capital punishment is morally wrong, and Montreal is south of New York.' I would classify that sentence as ethical: anyone who assents to the sentence takes on two commitments, one of them ethical. In spite of that commitment, though, Karmo's taxonomy has the sentence come out non-ethical, since the sentence is (actually) false regardless of what the correct ethical standard is. But I'm inclined to let that peculiar result pass, too.

Karmo's account has two further, more important, peculiarities.[9] According to the account, any sentence S is ethical provided at least two ethical standards differ on the truth-value of S. But ethical standards are themselves defined — 'completely determined' — by their assignment of truth-values to all the *uncontroversially* ethical sentences. If ethical standards are nothing more or less than complete (and consistent) assignments of truth-values to all of the uncontroversially ethical sentences, then they need not say anything about the truth-values of sentences that are merely ethical without being uncontroversially so. But, for Karmo, a sentence counts as ethical — never mind uncontroversially ethical — if and only if two ethical standards differ on its truth-value, and thus he ends up assimilating the class of ethical sen-

---

8 Karmo says nothing to rule out sentences having both of those properties, but it seems to me he has to rule them out, for reasons I give in note 24. In any event, my argument requires that I deny the existence of any such sentences — a denial which, although contentious, is dialectically legitimate in that context, as I argue below.

9 I thank an anonymous referee from this journal for suggesting that I confront these peculiar aspects of Karmo's account.

tences to the class of uncontroversially ethical sentences: a sentence will turn out to be ethical if and only if it is uncontroversially so.

His account thus risks giving away the game. First, if only uncontroversially ethical sentences are ethical at all, then the anti-autonomist's putative 'is'-'ought' derivation, such as the derivation I give below, will contain only non-ethical premises: its premises will not, of course, be *uncontroversially* ethical sentences — the anti-autonomist, one party to the controversy, denies that they are ethical — and so, on Karmo's account, they will not be ethical sentences at all. Second, some of Karmo's remarks suggest that every uncontroversially ethical sentence is non-contingently ethical, ethical in every possible world. Although he does not explicitly say that 'the uncontroversially ethical sentences' rigidly designates the set I have labeled 'U,' he does classify the sentence 'It ought to be the case that all New Zealanders are shot' as 'ethical in every possible world' (254), and that sentence is also one he gives as an example of the uncontroversially ethical. But if 'ethical' turns out to mean 'uncontroversially ethical,' and if 'uncontroversially ethical' turns out to mean 'noncontingently ethical,' then there will be no room in Karmo's taxonomy for contingently ethical sentences after all. As damaging as these features of his account appear to be, however, it would be too quick to dismiss his account on the basis of them. The criticisms that follow, then, will not turn on his having assimilated the ethical to the uncontroversially ethical.

III. Karmo's Reasoning

Relying on his taxonomy, Karmo gives the following reasoning for the claim that any derivation of an ethical conclusion from entirely non-ethical premises must be unsound. Suppose there is a world w in which a true sentence $S(n + 1)$ is both ethical and soundly derivable from, because it is entailed by, the true sentences S1 through Sn.

> [S]ince all of S1, ..., Sn are true in w — that is, are true with respect to the correct ethical standard E — S(n + 1) is true in w with respect to E. If S(n + 1) is to be ethical in w, then there must be some alternative ethical standard E' with respect to which S(n + 1) is false in w. Since replacing E with E' changes the truth value of S(n + 1) in w, and since S(n + 1) is entailed by S1, ..., Sn, replacing E with E' changes the truth value in w of at least one of S1, ..., Sn; so at least one of S1, ..., Sn is ethical in w [256-7].

Thus, he concludes, any derivation of an ethical conclusion is either unsound or else infected with at least one ethical premise.

IV. A Counterexample

Karmo's reasoning runs aground, however, for his taxonomy fails to reckon with a certain kind of sentence which, given non-question-begging assumptions, can be shown to be, when standardly construed, non-ethical in at least one world even though it satisfies his conditions for being an ethical sentence in every world. The sort of problem I will point out is hardly unique to Karmo's taxonomy; it confronts anyone else who argues that sound 'is'-'ought' derivations are impossible. Such derivations are indeed possible, and the following derivation, I will argue, is one example:

(B1)   Some ethical sentences, standardly construed, are true.

(B2)   Either no ethical sentence, standardly construed, is true, or torturing babies just for fun is morally wrong.

Therefore:

(B3)   Torturing babies just for fun is morally wrong.[10]

This derivation is obviously valid, since it relies on nothing more than disjunctive syllogism.[11] The conclusion, B3, is an ethical sentence if anything is, and on Karmo's criteria it counts as an ethical sentence in every world: in any world at all, B3 (taken as having its standard meaning) gets assigned 'true' by at least one ethical standard and 'false' by at least one other. (The latter ethical standard may be outrageous, but Karmo seems to allow that there are ethical standards to which no one

---

10   From B3 we can, I take it, deduce an explicit 'ought'-conclusion, e.g., the conclusion 'One ought not torture babies just for fun.' To those who object that B3 isn't explicitly an 'ought'-conclusion and that it doesn't straightforwardly entail one either, I offer a modified version of my derivation that produces an explicit 'ought'-conclusion without sacrificing plausibility in the premises:
 (B1) Some ethical sentences, standardly construed, are true.
 (B2*) Either no ethical sentence, standardly construed, is true, or one ought not torture babies just for fun.
 ☐ Therefore: (B3*) One ought not torture babies just for fun.

11   The logical form of the derivation is 'φ; ~φ v ψ; therefore, ψ.' So-called 'relevance' logicians notoriously contest the validity of disjunctive syllogism, but even a relevance logician (who doubts that disjunctive syllogism is an exceptionless rule) should see nothing wrong, as regards relevance, in my particular *use* of disjunctive syllogism. More to the present point, I doubt that anyone believes in the logical autonomy of ethics principally because she is suspicious of disjunctive syllogism.

would seriously subscribe.) Thus, on Karmo's account, the premises cannot both be true unless at least one of them is an ethical sentence.

But neither premise is an ethical sentence. Taking the premises in reverse order, I can show that B2 is (at least possibly and probably in fact) a non-ethical sentence by first proving a lemma and then conjoining the lemma with an obvious truth. Let 'ethical nihilism' (hereafter, 'EN') stand for the first disjunct in B2 — the thesis that no ethical sentence, standardly construed, is true. Karmo explicitly assumes that the parties to the debate over the autonomy of ethics are construing sentences as having their standard meanings (255), a sensible and crucial assumption that I too am making. He wants the assumption to be understood even where it is not otherwise explicit (ibid.), and so do I. Consider, then:

> *Lemma*: B2 and EN are both true in at least one world.

> *Proof*:   (a) $\lozenge$EN   [assumption]

>   (b) $\square$(EN $\supset$ B2)

>   (c) $(\lozenge\varphi \,\&\, \square(\varphi \supset \psi)) \supset \lozenge(\varphi \,\&\, \psi)$

> Therefore: (d) $\lozenge$(EN & B2)

Premise (a) assumes the logical possibility of EN — in other words, the logical possibility that every standardly construed ethical sentence is false.[12] While I think that EN is false (otherwise I wouldn't consider a sound derivation of an ethical sentence feasible), I will assume that EN is logically possible. I recognize the contentiousness of assuming the contingent falsity of EN, but I hasten to point out that the assumption begs no relevant questions in my dispute with Karmo and his fellow ethical autonomists. The most interesting arguments for the logical

---

12 Those attracted to ethical *noncognitivism* will object to my rephrasing EN in this way. On the noncognitivist view, an ethical sentence, such as B3, can fail to be true without thereby being false, since on that view ethical sentences typically lack either of the two classical truth-values. It is obvious, however, from Karmo's talk of true and false ethical sentences that he means to set aside noncognitivism, at least for the purpose of demonstrating the logical autonomy of ethics. Perhaps setting it aside is one intended effect of his stipulation that we construe sentences, including ethical sentences, 'with their standard meanings' (255): it seems fair to say that ethical noncognitivism construes the meanings of ethical sentences in a nonstandard way or, perhaps, that noncognitivism considers the standard way of construing ethical sentences (i.e., as typically true or false) not to be the best way of construing them. In any event, I join Karmo here in setting noncognitivism aside.

autonomy of ethics, including Karmo's, do not conclude that sound derivations of the ethical from the non-ethical are impossible on the grounds that every standardly construed ethical sentence is false or that, standardly construed, some ethical sentences are noncontingently true. On the contrary, the most interesting arguments for autonomy conclude that sound derivations of the ethical from the non-ethical are impossible even if some standardly construed ethical sentences are true and whether or not any such sentences are noncontingently true. It is these arguments that I am opposing here. To put it another way, if EN is true, then there is no ethical sentence that can serve as the conclusion of a sound derivation, in which case the ethical autonomist wins the debate by uninteresting default: because there are only false ones, ethical sentences turn out not to be soundly derivable from *any* kind of sentences. So naturally I have to assume the falsity of EN; what is more, I have to assume its contingent falsity. Again, however, no one concerned to argue for ethical autonomism in an interesting way will deny me those assumptions.

According to premise (b), EN entails — i.e., strictly implies — B2: EN materially implies B2 in every possible world. Assuming, with Karmo, that entailment is a relation between sentences, premise (b) is correct: B2 is just the disjunction of EN and B3, and '$\Box(\varphi \supset (\varphi \vee \psi))$' is a logical truth for any sentences φ and ψ. So EN entails B2.

Premise (c), expressed here as a schema, is a theorem of every standard system of modal logic: if there is a world in which φ is true, but no world in which φ is true while ψ is not, then there is a world in which both φ and ψ are true. The conclusion, (d), follows straightforwardly from the premises, although for brevity I have left out the subconclusions.

So B2 and EN are both true in at least one world. But it follows from the express terms of EN ('No ethical sentence…is true') that

(C)   No sentence is both an ethical sentence and a true sentence in any world in which EN is true.

Conjoining the lemma and principle C, we get the result that B2 is non-ethical in at least one world: there is a world in which both B2 and EN are true, but that can be so only if B2 is non-ethical in that world.

The fact that B2 is non-ethical in at least one world immediately falsifies Karmo's taxonomy of ethical sentences, for his taxonomy classifies B2 as an ethical sentence in every world. Recall that B3 is a noncontingently ethical sentence, since in any world at all B3 gets assigned 'true' by some ethical standard and 'false' by some other: in any world at all, there is an ethical standard E that forbids torturing babies just for fun and another (admittedly bizarre) standard E' that explicitly

permits such torture.[13] But B2 is the disjunction of EN and B3, and in any world at all E assigns B2's second disjunct (and thus B2 itself) the value 'true' while E' assigns both of B2's disjuncts (and thus B2 itself) the value 'false.'

Surely, moreover, there is such a thing as the 'nihilistic ethical standard,' the standard that assigns 'false' to every member of the set U of uncontroversially ethical sentences.[14] The nihilistic ethical standard is allowed for by Karmo's definition of an ethical standard as a complete (and consistent) assignment of truth-values to the members of U: why couldn't that assignment give the value 'false' to all such sentences? One might reply that some pairs of ethical sentences, such as

(B3) Torturing babies just for fun is morally wrong

and

(D) Torturing babies just for fun is not morally wrong,

are logical contradictories, in which case no consistent ethical standard could assign them both the value 'false.' But this reply equivocates on the interpretation of 'not morally wrong' in D. If we read 'not morally wrong' as 'morally permissible' (i.e., having the moral property of permissibility), then both B3 and D are ethical sentences, but they are not contradictories: since, I am assuming, EN is logically possible, there is a possible world in which B3 is false and in which D, read as attributing moral permissibility to baby torture, is also false. If, on the other hand, we read D as a wide-scope negation — i.e., as ~B3 — then of course B3 and D are contradictories, but D is not then an ethical sentence, for reasons that will emerge shortly.

Thus, given the logical possibility of EN, there seems to be no reason to deny the existence of the nihilistically ethical standard. But, if so, Karmo's taxonomy again makes B2 noncontingently ethical. In any world at all, the nihilistic standard assigns B2's nihilistic first disjunct

---

13 Surely we do not require that any genuine ethical standard be *defensible*; otherwise, the phrase 'defensible ethical standard' would be redundant (which it isn't), and 'indefensible ethical standard' would be a contradiction in terms (which it isn't). At any rate, Karmo does not seem to require it; one of the ethical standards he discusses requires the shooting of all New Zealanders (254-6).

14 A remark of Humberstone's suggests that he allows for the nihilistic ethical standard: he mentions an ethical standard 'according to which killing is not wrong (perhaps because nothing is)' ('A Study,' 157).

(and thus B2 itself) the value 'true' while a non-nihilistic standard that permits torturing babies just for fun assigns both of B2's disjuncts (and thus B2 itself) the value 'false.'

Worse yet, given the existence of the nihilistic ethical standard, Karmo's taxonomy makes thesis EN itself an ethical sentence in every world: EN gets assigned 'true' by the nihilistic standard and 'false' by every non-nihilistic standard. But, again on the assumption that EN is logically possible, EN is a non-ethical sentence in at least one world and, very likely, in every world. Assuming it is logically possible, EN is a non-ethical sentence in at least one world, since if it were an ethical sentence in every world it would be true in no world. Moreover, it seems likely that, regardless of its truth-value, EN has one status or the other, ethical or else non-ethical, across all worlds. Unlike premise A1, 'Everything that Alfie says is true,' whose propositional content (when A1 is standardly construed) varies with the content of Alfie's utterances, EN is a negated existential quantification whose propositional content (when EN is standardly construed) does not vary across worlds (although its truth-value does); also in contrast to A1, in accepting EN one does not run the risk of committing oneself to the truth of some uncontroversially ethical sentence. Therefore, EN's status as ethical or non-ethical should not vary across worlds. If so, then EN is not just possibly but necessarily a non-ethical sentence.

Thus, Karmo's taxonomy incorrectly classifies both B2 and EN as noncontingently ethical. B2 and EN are at least possibly non-ethical, and EN is very likely non-ethical in every world. In both cases, we have reason to reject Karmo's method of classifying ethical sentences.

Given my ambitions, though, I need to show more. In order to maintain that I have given a sound 'is'-'ought' derivation, I need to show that premise B2 is not just possibly but actually non-ethical. The only way I can see of doing so is to rebut the best available version of the contrary view: the version according to which, if premise B1 is in fact true, then premise B2 is in fact ethical.[15] According to that version, B2 is a non-ethi-

---

15   I thank David Robb for pressing this version of the contrary view. Some of Karmo's remarks hint at something like this version. In the context of my disjunctive premise B2, consider this remark, applied to a different disjunctive premise: '[T]he disjunctive premiss is ethical in w. (The disjunctive premiss has a false first disjunct in w. Consequently, its truth or falsity in w turns on [the truth of its second disjunct and thus on] the question whether the correct ethical standard prescribes or, on the contrary, refrains from prescribing the exceptionless shooting of New Zealanders)' (256). Karmo's problem (and the reason he cannot consistently accept this version of the contrary view) is that he does not reckon with a disjunctive premise one of whose disjuncts is EN itself.

cal sentence (and a true one) in any world in which its first disjunct, EN, is true, and B2 is an ethical sentence (whether or not a true one) in any world in which EN is false. The underlying reasoning is this: in a world in which EN is true, the truth of B2 does not turn on the truth of any non-nihilistic ethical standard, and so B2 is non-ethical in such a world; in a world in which EN is false, B2's truth does turn on the truth of some non-nihilistic ethical standard, and so B2 is ethical in such a world. If this version of the contrary view is correct, then I cannot maintain both that B1 (i.e., the negation of EN) is in fact true, as it must be if my derivation is sound, and that B2 is in fact non-ethical, as it must be if my derivation lacks any ethical premises.

Notice, first of all, that Karmo gets no help from this version of the contrary view, because he cannot consistently accept it. His taxonomy, again, even though it makes some sentences only contingently ethical, makes B2 noncontingently ethical: in any world there is an ethical standard E, whether or not correct, that forbids torturing babies just for fun and another standard E', whether or not correct, that permits it. In a world in which EN is true, both E and E' are incorrect standards, but, on Karmo's taxonomy, in order for B2 to be an ethical sentence in a world it suffices that E and E' differ on the truth-value of B2 in that world, even if neither E nor E' is correct in that world. Thus, for Karmo, B2 is ethical even in ethically nihilistic worlds — those worlds in which, according to the contrary view I'm now examining, B2 is non-ethical.

But how do I reply to the contrary view if it relies on a taxonomy of ethical sentences different from Karmo's? That depends on the taxonomy. The contrary view seems to rely on a taxonomy according to which

(F) φ is an ethical sentence in a world w if and only if EN is false in w and at least two ethical standards differ on the truth-value of φ in w.

Principle F, however, has the unwelcome consequence that B3 is an ethical sentence only in worlds in which EN is false. But surely B3 is an ethical sentence — though not, of course, a true sentence — even in worlds in which EN is true. Suppose, instead, that this version of the contrary view rests on a logically weaker taxonomic condition that is merely sufficient for a sentence's being ethical in a world:

(F*) φ is an ethical sentence in a world w if (but not only if) EN is false in w and at least two ethical standards differ on the truth-value of φ in w.

F* avoids the problem encountered by F, but it runs into two others. Given the existence of the nihilistic ethical standard, F* entails that EN

itself is an ethical sentence in any world in which EN is false: in any world *w*, including any world in which EN is false, the nihilistic ethical standard assigns EN the value 'true' while any non-nihilistic standard assigns it 'false.' F*, then, has EN come out ethical in any world in which EN is false, a result that creates two problems.

First, if EN is an ethical sentence in any world in which it is false, then whenever EN is false it is false *by its own lights*: whenever EN is false, it is exactly the kind of sentence that EN *says* is false. Surely there are other ways for EN to be false; surely there are worlds in which EN is false without belonging to the very class of sentences that EN itself says are false, but any such worlds are ruled out by F*.

The second problem is worse. As I have already argued, EN is non-contingently non-ethical: because it is logically possible EN is non-ethical in at least one world, and its status as ethical or non-ethical doesn't vary across worlds. If so, then F* entails that EN is necessarily true, i.e., false in all the worlds (namely, none) in which EN is an ethical sentence.[16] But the consequence that EN is necessarily true contradicts my assumption that EN is contingently false, and, more importantly, it is a consequence that the defender of F* will not want to accept. If EN is necessarily true, then F* vacuously entails that *any* sentence is an ethical sentence, since the antecedent 'if EN is false' becomes impossible to satisfy. If, on the other hand, we disallow vacuous entailments and apply F* only to worlds in which EN is false, then F* will never apply at all. In either case, F* cannot form the basis of a sound taxonomy of ethical sentences.

Thus, principles F and F* both have unacceptable consequences. But one or the other of them seems presupposed by the best version of what I've been calling 'the contrary view,' the view according to which premise B2 is in fact an ethical sentence if premise B1 is in fact true. I consider myself to have rebutted, then, the best version of the contrary view and, with it, the contrary view itself. So I tentatively conclude that B2 is not just possibly but actually non-ethical regardless of the truth of B1. I'll give one further reason for that conclusion in the course of discussing B1.

The classification of B1 is even harder to nail down. Given the existence of the nihilistic ethical standard, Karmo's taxonomy classifies B1 as an ethical sentence, but that is hardly a threat to my 'is'-'ought' derivation, since I have shown Karmo's taxonomy to have unacceptable consequences.

---

16 F* has the consequence $\Box(\sim EN \supset (EN \text{ is an ethical sentence}))$, which implies $\Box EN$ when conjoined with the result, $\Box\sim(EN \text{ is an ethical sentence})$, for which I argued earlier.

Karmo observes that his taxonomy has what he calls an 'appealing feature': 'if it makes a sentence S ethical at a world w, then it makes the negation of S ethical at w also' (257 n. 7). We have seen, however, that this feature of his account is entirely *un*appealing. Sentence B3, for example, is a noncontingently ethical sentence on Karmo's view as well as my own. However, by an analogue of the lemma I proved earlier, the negation of B3 — i.e., 'It is not the case that torturing babies just for fun is morally wrong' — is non-ethical in at least one world. Since EN is logically possible, and since EN necessarily implies ~B3, there is a world in which EN and ~B3 are both true. According to principle C, however, no sentence is both true and ethical in a world in which EN is true. So ~B3 is provably non-ethical in at least one world even though its negation, B3, is ethical in every world. Because Karmo's taxonomy is thus flawed, I won't use it here to take the easy way out, which would be to argue that, since the negation of B1 — i.e., EN — is non-ethical in every world, B1 itself must be non-ethical as well.[17]

Fortunately, though, there are other reasons for classifying B1 as non-ethical. Our partitioning of ethical and non-ethical sentences should, among other things, capture the distinction between sentences that are merely *about* morality (and thus aren't ethical sentences) and sentences that are genuinely *of*, or 'belong to,' morality (and thus are ethical sentences). Contrast, for instance, the conclusion of my derivation, B3, with the sentence

(G) Morality often goes by the label 'ethics.'

While, intuitively, B3 is a sentence *of* morality, G is a sentence merely *about* morality. One way of capturing this distinction is to require that any genuinely ethical sentence entail that a given moral property (such as moral rightness or wrongness) is possessed by at least one object (a person, action, event, institution, or whatever). Thus:

(H) φ is an ethical sentence only if it is impossible for φ, standardly construed, to be true unless a given moral property (such as moral rightness or wrongness) is possessed by at least one object.[18]

---

17 As recently as 'A Study,' Humberstone explicitly endorses Karmo's false conclusion (which Humberstone calls 'a pleasant feature') that 'the class of statements ethical (or, equivalently, non-ethical) in a world is closed under negation' ('A Study,' 152).

18 To put it symbolically, φ is an ethical sentence only if $(\exists x)\Box(\varphi \supset (Mx \ \& \ (\exists y)(Hyx)))$, where "$Mx$" abbreviates "$x$ is a moral property," and "Hyx" abbreviates "y possesses (property) x." The variable x ranges over moral properties such as rightness,

Admittedly, H is not a *definition* of 'ethical sentence.' It is just a necessary condition that any ethical sentence must satisfy and not a sufficient condition that makes ethical any sentence satisfying it. Moreover, in order to avoid an infinite regress, we may have to take something, perhaps the notion of a moral property, as primitive, as undefined, but then any (finite) conceptual analysis must take something as primitive. Given those caveats, however, I think that H will get us what we want: it leaves B3 with a chance of being ethical, since B3 satisfies the condition, but it makes each of EN, B1, and B2 non-ethical. Neither EN nor B1 nor B2 entails that a given moral property, such as moral rightness or wrongness, is possessed by some object or other: EN manifestly does not; B2 does not, since it can be true in virtue of the truth of EN; and B1 does not, since to say that B1 is true is not to commit oneself to the instantiation of a *given* moral property (only, at most, to the instantiation of some moral property or other).

Given the distinction between sentences of and sentences merely about morality, this is perhaps the best place to address one final objection to my claim that premise B2 is actually non-ethical.[19] According to the objection, there are two intuitively compelling reasons why B2 is actually an ethical sentence: (i) the truth-value of B2 depends on what the correct ethical standard is (i.e., whether that standard is the nihilistic one, or instead a non-nihilistic standard forbidding baby-torture, or instead a non-nihilistic standard permitting baby-torture); and (ii) B2 tells us about the moral structure of the world. As a matter of fact, though, neither reason is compelling. Reason (i), although it is a true statement, is just the reason given by Karmo's taxonomy, which taxonomy demonstrably fails to demarcate the ethical from the non-ethical. Reason (ii), on the other hand, would be a good reason if only it were true, but (ii) is false: B2 does not tell us anything about the moral structure of the world.

Perhaps an analogy can show why it does not. Let 'astrological nihilism' stand for the thesis that no distinctively astrological sentence, standardly construed, is true. I myself am an astrological nihilist; I think

---

wrongness, obligatoriness, permissibility, justice, injustice, and so on; *y* ranges over such objects as individual persons, action-types (such as lying or killing), action-tokens (such as particular lies or killings), institutions, practices, customs, and the like. This formula specifies a necessary condition and not, of course, a definition, but that way it avoids defining every impossible sentence, including "2+2=5," as an ethical sentence.

19 The objection is a familiar one, but I owe it in this particular form to an anonymous referee from this journal.

that there are plenty of true sentences about astrology but, as a matter of contingent fact, no true ones that belong to, or are distinctively of, astrology. Now consider the sentence

> (J) Either no distinctively astrological sentence, standardly con-
> strued, is true, or the Sun sign under which one is born influences
> one's personality.

As an astrological nihilist, I believe that J is true, because I believe its first disjunct is true. Do I think that J tells us anything about the 'astrological structure' of the world? No; I don't think there is any such structure in the first place for J to tell us anything about. Instead, J tells us something about the 'structure of astrology,' about the way the astrology game is played: namely, that among the best candidates for astrological truth, if there were any such truths, is the sentence 'The Sun sign under which one is born influences one's personality'; among distinctively astrological sentences, that sentence is logically so weak that it would be true if any astrological sentence were true. To assert J is to assert something about astrology, not to assert something astrological; the truth of J is entirely compatible with, because it is entailed by, astrological nihilism. But J is, of course, analogous to B2 in all the relevant respects: B2 tells us nothing about the moral structure of the world, because B2 is compatible with EN, the thesis that the world has no moral structure — no moral facts, no moral properties — for us to be told anything about.

If all of the foregoing is right, then the ethical conclusion B3 follows validly from the non-ethical premises B1 and B2, in which case Karmo's thesis commits him to the claim that at least one of those premises is false. But because his original argument relies on a taxonomy that simply gets it wrong, his original argument for the falsity of one or both of B1 and B2 won't work. We need, then, some other argument for the conclusion that the conjunction of B1 and B2 must be false, and Karmo provides none. It is hard to see how one could argue that the conjunction *must* be false, since it seems hard enough to argue that the conjunction is in fact false: B1 and B2 are surely not incompatible, and each of them strikes at least me as plausible in its own right.[20] Indeed, I have deliberately

---

20 The second premise of my derivation becomes even more plausible if we replace it
    with
        (B2**) Either no ethical sentence, standardly construed, is true, or torturing
        babies just for fun is *presumptively* morally wrong,
    since even those hedonistic utilitarians who allow for 'utility monsters' can accept
    B2**.

worded B1 and B2 so that they are logical subcontraries: compatible
sentences not both of which can be false. The less plausible you consider
one of those two premises, the more plausible you must consider the
other; you cannot consider them both implausible.[21] Granted, again, the
ethical nihilist considers their conjunction false because she considers B1
false; but, as I said before, the ethical nihilist regards the autonomism
debate as uninteresting — she thinks that the non-existence of sound
'is'-'ought' derivations is already established by the truth of nihilism.

So Karmo's argument no longer proves the impossibility of a sound
'is'-'ought' derivation. Moreover, the title of his article suggests that any
valid 'is'-'ought' derivation will be trivial, a suggestion I also promised
to rebut.[22] My derivation is nontrivial both logically and ethically. It is
logically nontrivial unless somehow every application of disjunctive
syllogism is logically trivial. I do not maintain that B3 is noncontingently
true and so automatically entailed by any premises.[23] On the contrary,
in order to prove the lemma, I have to assume that *no* ethical sentence is
noncontingently true.[24] Nor do I maintain that B1 or B2 is impossible or
that they are incompatible; on the contrary, as I've said, I think they are

---

21 Regarding B1 (i.e., ~EN) as (probably) false commits one to the (probable) truth of
    B2 (i.e., EN v B3), and regarding B2 as (probably) false commits one to the (probable)
    truth of (~EN & ~B3) and thus, of course, to the (probable) truth of B1. By contrast,
    however, regarding B1 as (probably) true does not commit one to the (probable)
    falsity of B2, and regarding B2 as (probably) true does not commit one to the
    (probable) falsity of B1.

22 I say 'suggests' because Karmo does not *argue* for the triviality of all valid 'is'-'ought'
    derivations.

23 On the classical conception of validity (according to which a derivation is valid if
    and only if it is impossible for its premises all to be true and its conclusion false): (i)
    any derivation with logically impossible or mutually inconsistent premises is
    classically valid, no matter what its conclusion; (ii) any derivation whose conclusion
    is noncontingently true is classically valid, no matter what its premises (and
    classically sound provided only that its premises are true). My derivation is classi-
    cally valid, but not in either of those trivial ways.

24 For this reason, and in order to steer clear of a logically trivial derivation of the
    ethical from the non-ethical, I have not used the following, even quicker, way of
    rebutting Karmo: if, as some philosophers maintain (e.g., Judith Jarvis Thomson,
    'The No Reason Thesis,' *Social Philosophy & Policy* 7 [1989]: 1-21; 9), there are
    noncontingently ethical, noncontingently true sentences, then those ethical sen-
    tences will be trivially derivable from any sentences at all — and soundly (though
    often only trivially) derivable from any true sentences at all, including any true
    non-ethical sentences.

both quite plausible. In any event, I do not hold that B1 and B2 trivially entail B3 in virtue of entailing any conclusion at all.

The derivation is also ethically nontrivial. One might well use it, as I have used it, to try to convince a non-nihilistic ethical skeptic to abandon her unstable epistemological position.[25] The non-nihilistic ethical skeptic allows that B1 may be true, but she characteristically insists that no one knows the truth of B3 or, for that matter, any other ethical sentence. My (knowably valid) derivation can be used to persuade her of the implausibility of that view: since neither premise B1 nor premise B2 is an ethical sentence, her ethical skepticism, as such, gives her no reason to deny anyone knowledge of those premises; but then, assuming that she accepts the closure of knowledge under known logical entailment, she has no reason to deny anyone knowledge of the ethical conclusion, B3 — and there goes her ethical skepticism. I think it is fair to classify as ethically nontrivial any derivation that has the prospect of overturning non-nihilistic ethical skepticism.

If I am right, then I have given a logically and ethically nontrivial, potentially (indeed, I think, probably) sound derivation of an ethical conclusion from two non-ethical premises. If it is even possible that I have done so, then Karmo's thesis cannot be correct, and ethical autonomism more generally is in trouble.[26]