

THE LOGICAL POSSIBILITY OF MORAL DILEMMAS IN EXPRESSIVIST SEMANTICS: A CASE STUDY

Ryo Tanaka¹

¹University of Tokyo, Japan

Original scientific paper – Received: 09/09/2023 Accepted: 26/02/2024

ABSTRACT

In this paper, using Mark Schroeder's (2008a) expressivist semantic framework for normative language as a case study, I will identify difficulties that even an expressivist semantic theory capable of addressing the Frege-Geach problem will encounter in handling the logical possibility of moral dilemmas. To this end, I will draw on a classical puzzle formulated by McConnell (1978) that the logical possibility of moral dilemmas conflicts with some of the *prima facie* plausible axioms of the standard deontic logic, which include *obligation implies permission*. On the tentative assumption that proponents of ethical expressivism should be generally committed to securing the logical possibility of moral dilemmas in their semantic theories, I will explore whether and how expressivists can successfully invalidate *obligation implies permission* within the framework developed by Schroeder. The case study eventually reveals that this can indeed be a hard task for expressivists. Generalizing from the case study, I will suggest that the source of the difficulty ultimately lies in the mentalist assumption of the expressivist semantic project that the logico-semantic relations exhibited by normative sentences should be modeled in terms of the psychological attitudes that speakers express by uttering them. My final goal will be to show that the difficulty expressivists face in dealing with the logical possibility of moral dilemmas is a reflection of the more general problem that their commitment to the mentalist assumption prevents them from flexibly adopting or dropping axioms in their semantic theories to get the right technical results.

Keywords: expressivism; moral dilemmas; metaethics; semantics; deontic logic.

1. Introduction

In this paper, using Mark Schroeder's (2008a) expressivist semantic framework for normative language as a case study, I will identify difficulties that even an expressivist semantic theory capable of addressing the Frege-Geach problem will encounter in handling the logical possibility of moral dilemmas. To this end, I will draw on a classical puzzle formulated by McConnell (1978) that the logical possibility of moral dilemmas conflicts with some of the *prima facie* plausible axioms of the standard deontic logic, which include *obligation implies permission*. On the tentative assumption that proponents of ethical expressivism should be generally committed to securing the logical possibility of moral dilemmas in their semantic theories, I will explore whether and how expressivists can successfully invalidate *obligation implies permission* within the framework developed by Schroeder. The case study eventually reveals that this can indeed be a hard task for expressivists. Generalizing from the case study, I will suggest that the source of the difficulty ultimately lies in the mentalist assumption of the expressivist semantic project that the logico-semantic relations exhibited by normative sentences should be modeled in terms of the psychological attitudes that speakers express by uttering them. My final goal will be to show that the difficulty expressivists face in dealing with the logical possibility of moral dilemmas reflects the more general problem that their commitment to the mentalist assumption prevents them from flexibly adopting or dropping axioms in their semantic theories to get the right technical results.

In the remainder of the introduction, I will address three preliminary issues. First, I will introduce the puzzle concerning the logical possibility of moral dilemmas that I will discuss in this paper. Second, I will explain and briefly justify the tentative assumption of the paper that expressivists generally need to secure the logical possibility of moral dilemmas in their semantic theories. Lastly, I will explain why I specifically draw on Schroeder's framework to develop my discussion. The subsequent sections will be devoted to the case study: I will show how the problem concerning the logical possibility of moral dilemmas for ethical expressivism arises taking a specific shape in Schroeder's framework and explore how one can respond to it.

1.1 Ethical expressivism and the logical possibility of moral dilemmas

A moral dilemma is defined as a situation where incompatible courses of action A and B are both morally obligatory for an agent. The reality of such dilemmas in human life seems indubitable. One can easily imagine, and

often find oneself in, situations where general moral precepts such as “Do not lie”, “Do not steal”, “Help your family and friends”, may come to conflict with one another. One can see the reality of moral dilemmas most vividly in situations where one and the same moral precept seems to generate conflicting but equally strong demands. Consider the often-cited case of Sophie’s choice (Styron 1979): Sophie and her two children are at a Nazi concentration camp, and a guard tells Sophie that only one of her children will be allowed to live but the other will be killed. For each child, Sophie has an obligation to save him/her, but she cannot save both. In this case, it is implausible to think that Sophie can resolve the conflict by thinking that one of her obligations overrides the other, because there is no obvious reason why either of them should be stronger than the other. This and similar examples suggest that genuine, that is, irresolvable dilemmas are possible and often real.¹

As McConnell (1978, 2022) points out, however, the possibility of irresolvable moral dilemmas apparently conflicts with some of the *prima facie* plausible axioms of the standard deontic logic. (The presentation of the problem below follows McConnell (1978)). Crucially, the problem concerns the *logical* possibility of moral dilemmas—adopting the relevant axioms leads to the result that moral dilemmas are impossible as a matter of *logic and the meanings of the relevant normative expressions alone*. For the purpose of this paper, I will specifically focus on the two axioms, which are meant to capture the following *prima facie* plausible theses: (1) *permission can be defined in terms of obligation*, and (2) *obligation implies permission*.² Let OA stand for “A is obligatory” and PA stand for “A is permissible”. First, a moral dilemma is a situation where incompatible courses of action are both obligatory. To capture its troublesome nature, one can characterize a moral dilemma as a situation where the following holds:

$$(MD) OA \wedge O\sim A.$$

A moral dilemma is a situation where A is obligatory, but not doing A is obligatory as well because it is necessary for doing B, another obligatory action—e.g., saving one child’s life is obligatory for Sophie, but not doing so is obligatory as well because it is a necessary means for saving the life of the other child, which is another obligatory act for her. Second, (1) states that doing A is permissible if and only if it is not the case that not doing A is obligatory:

¹ For influential arguments for the logical possibility of moral dilemmas that invoke the notion of moral residue, see e.g., Marcus (1980), Tessman (2015), and Williams (1966).

² Brink (1994), for example, takes these as conceptual truths concerning the notions of obligation and permission.

$$(1) PA \Leftrightarrow \sim O\sim A.$$

Lastly, the symbolic representation of (2), *obligation implies permission*, is the following:

$$(2) OA \Rightarrow PA.$$

(1) and (2) entail $OA \Rightarrow \sim O\sim A$. Assuming the material conditional, $OA \Rightarrow \sim O\sim A$ is equivalent to $\sim(OA \wedge O\sim A)$. This, however, directly contradicts (MD). Hence, (1) and (2) are jointly inconsistent with (MD). Adopting the apparently intuitive theses (1) and (2) as axioms in one's theory thus rules out the logical possibility of moral dilemmas.

Since McConnell (1978) provided a formal presentation of the puzzle, many answers have been proposed in the literature.³ Those who think moral dilemmas should be at least logically possible seek ways to justify abandoning (at least) one of the axioms that give rise to the inconsistency, whereas opponents argue that any putative solution to the puzzle is bound to be *ad hoc*.⁴ (At this point, it may need to be noted that there are also other combinations of axioms that are known to be inconsistent with the logical possibility of moral dilemmas. I will come back to this point below.)

In this paper, for the sake of discussion, I tentatively assume that expressivists should generally side with the pro-dilemma view.⁵ A *prima facie* justification for this assumption stems from the fact that ethical expressivism is usually construed as a *semantic view*, that is, a view about *the meanings of normative sentences in a natural language*, such as English. (In the next section, I will elaborate on this point in more detail.) As such, proponents of ethical expressivism should be committed to formulating a theory that correctly reflects ordinary speakers' use of, and linguistic intuitions about, the target normative expressions—they constitute the data that should guide one's theory construction. Crucially, ordinary speakers' language use and linguistic intuitions seem to suggest

³ For a helpful survey, see McConnell (2022). For various responses to McConnell (1978), see e.g., Almeida (1990), Conee (1982), Goble (2009), Hansson (2019), Holbo (2002), Marcus (1980), McConnell (1978), Nair (2016), Sinnott-Armstrong (1988), Vallentyne (1989), and Zimmerman (1996). See e.g., Lemmon (1962) and van Fraassen (1973), for influential discussions of this topic that precede McConnell (1978).

⁴ Those who deny the logical possibility of moral dilemmas often point to the fact that historically influential philosophers such as Aquinas, Mill, Kant, Ross etc. seem to hold similar views. See Marcus (1980), McConnell (2022, 6), Sinnott-Armstrong (1988) for discussion.

⁵ As I will explain, the plausibility of the conclusions of the paper will not depend on the truth of this assumption. Certainly, all things considered, it might turn out that one's semantic theory should give up the possibility of moral dilemmas, rather than *obligation implies permission*. However, to avoid unnecessary complications, I will not address this point until Section 4.

that they take moral dilemmas to be at least conceptually possible. Ordinary speakers do often find themselves in dilemmatic situations and describe them as such, and, accordingly, they do not seem to take “A is obligatory for S, but not-A is also obligatory for S” as an utterly confused, non-sensical statement. To accommodate the data, one’s expressivist semantic account should not entail that moral dilemmas are impossible, *as a matter of logic and meaning alone*. When conjoined to the puzzle concerning the logical possibility of moral dilemmas presented above, this means that expressivists should construct their semantic theories so that *obligation implies permission* will not turn out to be formally valid (assuming that they do not want to reject the interdefinability of obligation and permission).⁶

To see the intuitive plausibility of this assumption, it may also be helpful to point out the fact that those who deny the possibility of moral dilemmas are usually motivated by substantive moral-theoretic concerns that do not necessarily coincide with ordinary speakers’ language use and linguistic intuitions (McConnell 2022, Sec. 3 and 4). For example, a Kantian theorist might argue that moral dilemmas should be impossible, because one’s (deontological) moral theory should be *uniquely action-guiding* in the sense that it will never prescribe incompatible actions for an agent in a given situation. To meet this requirement, one might propose that there should be some way of hierarchically structuring moral precepts so that irresolvable conflicts will never arise (for an influential critique of this idea, see e.g., Ross 1930, Ch. 2). Whether this line of reasoning is plausible or not, it clearly concerns a requirement that is to be imposed on one’s *substantive theory of morality*, not one’s *semantic account* of normative expressions in a natural language. Semantics must respect the fact that ordinary speakers often talk about moral dilemmas meaningfully and they do seem to take dilemmas to be at least conceptually possible. The assumption of this paper is that this constitutes a *prima facie* reason for expressivists to try to secure the logical possibility of moral dilemmas in their semantic accounts. Whether the possibility of moral dilemmas should be ultimately excluded in one’s substantive moral theory is a separate question, which I do not intend to address in the current paper.

⁶ One might question why expressivists need to reject *obligation implies permission* rather than the interdefinability of obligation and permission. In this paper, I will explore the former option mainly for pragmatic reasons (I leave open, but will not discuss in detail, the possibility of pursuing the other option). As I explain in the next paragraph, Schroeder himself discusses this problem by focusing on the question how one may (in)validate *obligation implies permission* in his framework. Schroeder does not pursue the alternative route since he thinks it “is an old observation that ‘permissible’, ‘impermissible’, ‘obligatory’, and ‘unobligatory’ can all be interdefined using negation” (Schroeder 2008a, 46). Furthermore, as I will explain in Section 3, rejecting the interdefinability of obligation and permission in Schroeder’s framework actually turns out to be more difficult compared to *obligation implies permission*.

In the rest of the discussion in this paper, as a case study, I will examine how this general problem concerning the logical possibility of moral dilemmas arises taking a specific shape in the expressivist semantic framework developed by Mark Schroeder (2008a) and explore how one might respond to it. There are two reasons why I specifically focus on Schroeder's expressivist semantics to explore this issue. The first is that, although ethical expressivism is one of the most discussed views in the literature on the semantics of normative language, apparently no proponents of expressivism have done a better job than Schroeder in actually constructing a semantic theory that seems to provide a promising response to the so-called Frege-Geach problem (Geach 1960, 1965). The Frege-Geach problem, in short, is a demand for expressivists to develop a *compositional* semantic theory that correctly captures logical-semantic relations between non-atomic, logically complex normative sentences (e.g., negations, disjunctions, conditionals, sentences with quantifications, etc.). In most expressivist proposals before Schroeder, the details of expressivist semantic theories were not fully spelled out, and it was even unclear if they could adequately explain such basic semantic phenomena as the logical inconsistency of "A is wrong" and "A is not wrong"—as I will explain in the next section, this is what Unwin (2001) calls the "negation problem" for expressivists, which is an instance of the Frege-Geach problem as applied to negation. An important contribution of Schroeder's work is that it identifies a *structural* requirement that any expressivist semantic account should meet to deal with the Frege-Geach problem.⁷ Exploring the issue of the logical possibility of moral dilemmas in his framework will provide a useful case study, because my initial goal is to show that an expressivist semantic theory that provides an adequate solution to the Frege-Geach problem does not necessarily succeed in securing the logical possibility of moral dilemmas as well. The second reason is purely pragmatic: at one place in his book, Schroeder (2008a, Ch. 5, Sec. 4) himself discusses the relevant thesis, *obligation implies permission*, in connection to the logical possibility of moral dilemmas. I will develop my own discussion by building on Schroeder's.⁸

⁷ For this reason, Schroeder's expressivist semantics is worth exploring also for those who are attracted to the recent movements of applying the expressivist idea to other types of discourse than the ethical. See e.g., Bar-On and Sias (2013) for discussion.

⁸ Also, it should be noted that the overall aim of Schroeder's book *Being For* is to illustrate the *costs* of ethical expressivism (see *Preface*). Schroeder's intention is to reveal the theoretical commitments that expressivists should make by actually developing a workable expressivist semantic theory on behalf of them. Throughout the book, he occasionally reminds readers to think about where to "get off the boat"—it is beyond the scope of the discussion in this paper to assess the overall plausibility of Schroeder's project construed as a *reductio* of ethical expressivism. See, however, Section 4 below for a brief discussion on the prospect of the expressivist semantic project in light of the result of the current paper.

Relatedly, as I noted earlier, there are also other combinations of intuitively plausible axioms that are known to be inconsistent with the logical possibility of moral dilemmas. For example, *obligation implies possibility* (i.e., *ought implies can*) and the principle of agglomeration, which states that *if A is obligatory and B is obligatory, then A and B is obligatory* (Williams 1966), are jointly inconsistent with the possibility of moral dilemmas (I will not provide a proof here—interested readers should consult surveys such as McConnell 2022, Sec. 4 and McNamara and Van De Putte 2023, Sec. 6.4.) If expressivists need to side with the pro-dilemma view, they will ultimately have to find ways to avoid such combinations as well. Due to the limit of space, I will not explore other combinations and focus only on *obligation implies permission*. In Section 4, I will briefly discuss how one might interpret the results of this paper in light of this broader point.

The rest of the discussion in this paper will proceed as follows. In Section 2, I will briefly review several basic features of ethical expressivism construed as a semantic project and introduce Schroeder's expressivist semantic theory against that general background. As noted, Schroeder's theory is mainly motivated as a response to the negation problem, which is a special instance of the Frege-Geach problem. The discussion may get technical in places, and to avoid unnecessary complications, in Section 2, I will *not* discuss how the technical tools developed by Schroeder connect with the issue of the logical possibility of moral dilemmas. I will turn to this issue in Section 3: I will explore how one might respond to the problem of securing the logical possibility of moral dilemmas within Schroeder's framework and make some general observations from the case study. In Section 4, I will conclude by articulating the moral of the case study in the most general terms.

The last caveat before the main discussion: in presenting Schroeder's expressivist semantics, I will, as Schroeder himself does, focus on the predicate "is wrong" as the main target of semantic analysis, instead of predicates such as "is obligatory" or "is permissible" that I used in introducing the issue of the logical possibility of moral dilemmas. As I will explain in Section 3, however, there is an easy and relatively uncontroversial way to translate claims that contain the predicate "is wrong" to those that only contain "is obligatory" and "is permissible" (for example, "A is obligatory" could be rephrased as "Not A is wrong"). Although this invites some complication, it does not pose any serious problem for the main discussion of this paper. Schroeder himself also notes that his discussion is applicable to expressivist views that take different normative predicates (e.g., "is rational", "is the thing to do", "ought", and

so on) as basic (Schroeder 2008a, 7, see, also, 39). I will come back to this point in Section 3.2.

2. Schroeder's expressivist semantics: A structural solution to the negation problem

Ethical expressivism is characterized by the idea that what one does when uttering a sentence with a normative predicate is to express one's non-cognitive attitude toward an object of evaluation. The basic expressivist idea can be traced to early non-cognitivist views proposed by Ayer (1936), Hare (1952), and Stevenson (1937); Blackburn (1984, 1988, 1998) and Gibbard (1990, 2003) are known for more systematic formulations of contemporary expressivist views (for a brief history of ethical expressivism, see Schroeder 2010, Chapter 4). On a simple expressivist account, for example, "Murder is wrong" might express a speaker's non-cognitive attitude of disapproval of murder. *Non-cognitive* attitudes contrast with cognitive attitudes (e.g., one's belief that murder has such and such properties) in that the former do not have truth-evaluable propositions as their contents. Importantly, as I mentioned in the previous section, expressivism is usually construed as a *semantic* view. To develop a semantic theory for descriptive language, one can fruitfully invoke the notion of truth-evaluable proposition. To develop a semantic theory for normative language, expressivists insist, a different approach is called for—the meanings of normative sentences are *not* truth-conditions. Instead, their meanings are the non-cognitive attitudes that speakers express by uttering them.⁹ Hereafter, I call this the *mentalist* assumption of the expressivist semantic project.¹⁰ It is the assumption that normative sentences must get their meanings from speakers' *mental states*, instead of propositions, which are usually taken to be abstract entities that exist independently of speakers' psychology.

On this construal, one major task for proponents of expressivism is to provide a systematic account of the logical and semantic features that normative sentences in our natural language seem to exhibit, without assuming the standard truth-functional compositional semantics developed primarily for descriptive language. Specifically, the mentalist assumption of the project requires that they should somehow model the logico-

⁹ It should be noted that some authors argue, against the orthodoxy, that expressivism need not be interpreted as a *semantic* thesis (Bar-On and Chrisman 2009, Bar-On et al. 2014). In Section 4 (in the last footnote), I will briefly discuss how one might interpret the results of this paper in connection with this kind of "neo-expressivist" positions.

¹⁰ This corresponds to what Sias (2024) calls "semantic ideationalism" in his survey entry on ethical expressivism.

semantic relations exhibited by normative sentences in terms of the psychological attitudes speakers express by uttering them. Logicians and formal semanticists can explain (among many other things) why a descriptive sentence, say, “Grass is green”, is inconsistent with “Grass is *not* green”, by appealing to the truth-functional definition of the meaning of “not” and the meaning (i.e., truth-condition) assigned to the original sentence. If expressivism is a view about meaning of normative language, it is expected that proponents of the view should be able to explain, in some parallel way, why “Murdering is wrong” is inconsistent with “Murdering is not wrong”. Informally, in the standard truth-conditional semantics, what negation does when applied to a descriptive sentence is to “flip” the truth-value assigned to the sentence. However, because expressivists understand the meanings of normative sentences in terms of non-cognitive attitudes they express instead of truth-conditions, what negation does when applied to normative sentences should be explained in a different way.

Of course, there is no principled reason why one ought to think that it is impossible to develop an expressivist semantic theory for normative language that meets this challenge.¹¹ Attempts have been made, most notably by Simon Blackburn and Allan Gibbard, at sketching outlines of compositional semantic theories for ethical language based on the expressivist assumptions (Blackburn 1984, 1988, 1998; Gibbard 1990, 2003). However, as Schroeder contends (2008a, *Preface*), there was no consensus whether they even succeeded at providing a plausible explanation of how “Murder is wrong” should be logically inconsistent with “Murder is not wrong”—this is known as “the negation problem” for expressivism (Unwin 2001). As mentioned above, the negation problem is an instance of the so-called Frege-Geach problem, which questions how proponents of expressivism (or non-cognitivism in general) could provide a compositional semantic account for non-atomic, logically complex normative sentences in such a way that the theory can correctly capture their logical-semantic relations. In Part II of the book, Schroeder sets out to develop his expressivist semantic account primarily by responding to the negation problem formulated by Unwin (see, also, Schroeder 2008b, 2008c). In the rest of this section, I will introduce Schroeder’s account by explaining how it is tailored to deal with the negation problem.

The negation problem, as formulated by Unwin, is this. Expressivists maintain that “Murdering is wrong” expresses one’s non-cognitive attitude toward murder—let us stipulate that it expresses the attitude, *disapproval of murder*. Then, what attitude should be assigned as the meaning of

¹¹ See, e.g., Hare (1970) for an expression of optimism about the prospect of non-cognitivism construed as a semantic project.

“Murdering is not wrong”, the sentence that should turn out to be logically inconsistent with the original sentence? At the first glance, *disapproval of not murdering* might seem to be a good candidate, because it seems logically inconsistent to disapprove of both ϕ -ing and not ϕ -ing. However, this cannot be right, because *disapproval of not murdering* should be, intuitively, the attitude that is to be expressed by “Not murdering is wrong” instead of “Murdering is not wrong”. These sentences clearly have different meanings, and no adequate semantic theory should conflate the meaning of one with that of the other. One might think that there should be some way of getting around the problem by inserting “not” in the right places in the attitudes expressed by the relevant sentences, but Unwin’s discussion shows that the problem cannot be solved so easily.

According to Schroeder (following Unwin), the negation problem arises from the “*insufficient structure*” (Schroeder 2008a, 57) in the attitudes expressed by, and thereby assigned as the meanings of, normative sentences. (Hereafter, all the references are to Schroeder (2008a) unless otherwise noted.) One can best see this point by looking at the following table (45; slightly modified from the original):¹²

- w Jon assents to “Murdering is wrong”.
- n1 Jon does not assent to “Murdering is wrong”.
- n2 Jon assents to “Murdering is not wrong”.
- n3 Jon assents to “Not murdering is wrong”.

- w* Jon disapproves of murdering.
- n1* Jon does not disapprove of murdering.
- n2* ???
- n3* Jon disapproves of not murdering.

The problem, in short, is that the account allows for too *few* ways to negate w*. There are only two places where one can insert “not” in *Jon disapproves of murdering* (which yield n1* or n3*), whereas there are three ways to negate w (n1, n2, and n3). Specifically, as I explained, expressivists need the attitude expressed by “Murdering is not wrong” (n2) to be inconsistent with *disapproval of murdering*, which is expressed by “Murdering is wrong” (w) —but, at the same time, the attitude in question cannot be *disapproval of not murdering*, because it should be assigned as the meaning of “Not murdering is wrong” (n3). Apparently, then, there seems to be no way of arriving at the correct semantic assignments for w,

¹² Schroeder, following Unwin, uses “Jon thinks that murdering is wrong” and so on in demonstrating the negation problem. Here and in the relevant places below I will use “Jon assents to ‘Murdering is wrong’” instead to highlight the fact that the problem primarily concerns which attitudes should be assigned as the meanings of normative *sentences*.

n_1 , n_2 , and n_3 , starting from the assumption that the meaning of “ ϕ -ing is wrong” is one’s *disapproval of ϕ -ing*.¹³

One way to avoid the problem might be to think that “Murdering is not wrong” should express a different kind of attitude than disapproval, such as one’s *tolerance of murdering*. This might allow expressivists to explain the inconsistency between “Murdering is wrong” and “Murdering is not wrong” by appealing to the stipulation that disapproval of ϕ -ing is inconsistent with tolerance of ϕ -ing (Blackburn 1988). However, Schroeder argues that this is a problematic move because it leaves completely unexplained why “two *distinct* and apparently *logically unrelated* attitudes [i.e., disapproval and tolerance] toward the *same* content” (48) can be logically inconsistent with one another. Schroeder contrasts this to the unproblematic kind of inconsistency that holds between two attitudes of the *same* kind toward inconsistent contents (ibid.). In Schroeder’s terminology, these are “inconsistency-transmitting attitudes”:

Inconsistency-transmitting attitudes: An attitude A is inconsistency-transmitting just in case two instances of A are inconsistent just in case their contents are inconsistent. (43)

Belief is a good example: believing that p is inconsistent with believing that not- p , *because* their contents, p and not- p , are logically inconsistent. In other words, in the case of belief, the inconsistency of the contents p and not- p *transmits* to one’s attitudes towards these contents. And insofar as the idea that belief is an inconsistency-transmitting attitude is generally accepted, there is no reason why expressivists cannot treat, say, disapproval as an inconsistency-transmitting attitude and assume that one’s *disapproval of murdering* is logically inconsistent with one’s *disapproval of not murdering*. On the other hand, Schroeder contends, it is not justified for expressivists to take it for granted that *disapproval of murdering* should be logically inconsistent with *tolerance of murdering*. On his view, this is a purely *ad hoc* solution to the negation problem, because it is a mere convenient stipulation that there should be non-cognitive mental attitudes of disapproval and tolerance such that they are completely distinct but nonetheless can be logically inconsistent with one another in some way. What makes this stipulation particularly problematic is the fact that, unlike

¹³ For more detailed presentations of the negation problem, see Unwin (2001) and Schroeder (2008a, Ch. 3, 2008b, 2008c).

inconsistency-transmitting attitudes, there are no undisputed good examples of attitudes that exhibit the desired feature (47-9).¹⁴

Taking stock: on Schroeder's view, the negation problem arises from the lack of structure in the attitudes (e.g., *disapproval of φ -ing*) that expressivists assign as the meanings of normative sentences. Furthermore, there is also the constraint that expressivists should not explain the inconsistency between normative sentences by stipulating the existence of multiple attitudes (e.g., disapproval and tolerance), each of which are primitive but nonetheless can be logically related.

Schroeder's main positive proposal defended in the book is that one can resolve the negation problem on behalf of expressivists by replacing the attitude of disapproval with a primitive inconsistency-transmitting non-cognitive attitude that he calls "*being for*" (58). Schroeder's overall strategy is to use this, and *only* this, attitude as the basic tool for constructing meanings for all normative sentences (hence, the title of the book, *Being For*). For the purposes of this paper, it would not be necessary to discuss Schroeder's exposition on the psychological nature of the attitude in question. The key point is that the attitude of being for creates the necessary *structure* that was missing in the expressivist semantic analysis that adopts the attitude of disapproval as the basic explanatory tool. On Schroeder's proposal, " φ -ing is wrong" expresses the attitude of *being for blaming for φ -ing*. Crucially, unlike *Jon disapproves of murdering*, there are three, instead of two, places to insert "not" in *Jon is for blaming for murdering*. The semantic analysis of w, n1, n2 and n3 that results from this proposal is shown in the following table (59):

- w Jon assents to "Murdering is wrong".
- n1 Jon does not assent to "Murdering is wrong".
- n2 Jon assents to "Murdering is not wrong".
- n3 Jon assents to "Not murdering is wrong".

- w** Jon is for blaming for murdering.
- n1** Jon is not for blaming for murdering.
- n2** Jon is for not blaming for murdering.
- n3** Jon is for blaming for not murdering.

Hereafter, following Schroeder's notation, I will abbreviate "being for blaming for φ -ing" as "FOR(blaming for φ -ing)". On Schroeder's proposal, one can explain the inconsistency between "Murdering is wrong"

¹⁴ However, for an important critique of Schroeder's argument summarized in this paragraph, see Baker and Woods (2015).

and “Murdering is not wrong” by the fact that the attitudes expressed by these sentences—i.e., FOR(blaming for murdering) and FOR(not blaming for murdering)—have inconsistent *contents*. On the assumption that the attitude of being for is (like belief) inconsistency-transmitting, these are inconsistent attitudes *because* they have inconsistent contents.

Thus, analyzing the meanings of w , $n1$, $n2$, and $n3$ in terms of being for attitudes provides expressivists with a systematic way of correctly capturing their logical relationships without making any controversial assumptions. More formally, the meanings of (i.e., the attitudes expressed by) any normative sentences that contain negation can be determined compositionally by applying the definition of negation, provided on p. 66:

(NEG) Where ‘A’ expresses FOR(α), ‘ \sim A’ expresses FOR($\sim\alpha$).¹⁵

From this definition, it follows that “Murdering is not wrong” expresses FOR(not blaming for murdering), that is, ($n2^{**}$). And, it is also natural to think that “Not murdering is wrong” expresses FOR(blaming for not murdering). This assignment of attitude is intuitive, and, more importantly, the assigned attitude is distinct from the one expressed by “Murdering is not wrong”. The analysis thus avoids conflating the meanings of “Not murdering is wrong” and “Murdering is not wrong”. This, Schroeder argues, resolves the negation problem on behalf of expressivists.

For the sake of discussion, I assume that Schroeder’s proposal summarized above provides a promising solution to the negation problem. Here, I want to highlight two features of Schroeder’s semantic framework that will be important for the purposes of the discussion below. Crucially, both derive from the fact that the negation problem is a *structural* problem and the solution requires adding the necessary structure to the attitudes assigned as the meanings of normative sentences. First, Schroeder’s discussion, if successful, implies that any proponent of expressivism should ultimately adopt a semantic theory that at least shares the basic structure with Schroeder’s account—that is, the structure that allows one to deal with the negation problem. As he puts it, adopting his semantic framework “isn’t just *a* way of making progress on the negation problem, for expressivists”—rather, it is “*the* expressivist solution to the negation problem” (61). Second, this need not mean, however, that expressivists should adopt the semantic theory that analyzes “Murdering is wrong” in terms of the attitude of being for blaming for murdering, *specifically*. Since the negation problem is a structural problem, any account that yields

¹⁵ In a similar fashion, Schroeder provides recursive definitions for conjunction and disjunction on page 66; he also defines entailment relationship between sentences on page 70.

sufficient structure should be able to deal with it, at least in principle. Schroeder himself notes in passing that he sticks with this specific analysis “just to fix examples” (58) and one could adopt an alternative stipulation that analyzes the meaning of “Murdering is wrong” as “being for *disapproving* of [murdering]” (ibid., emphasis mine), instead of being for *blaming* for murdering. Elsewhere, he also considers a proposal that “Murdering is wrong” expresses “being for avoiding murdering” (74). So, depending on one’s interests and pre-theoretical intuitions, which *specific* kind of attitude/act should be taken as the target of the being for attitude may vary, as long as it retains the structure necessary for dealing with the negation problem.

To summarize, the general conclusion of Schroeder’s discussion is that the attitude that is to be assigned as the meaning of “ ϕ -ing is wrong” should take the following form: *being for [one’s preferred unary expression] ϕ -ing*.¹⁶ This is the structural requirement that one’s semantic theory should meet to avoid the negation problem (and, more generally, to deal with the Frege-Geach problem). However, at the same time, this structural requirement does not entail any strong material restriction on which specific kind of attitude/act should go into the placeholder. In principle, any unary expression that takes a gerund (ϕ -ing) as the object will do, as long as it does not yield an obviously implausible meaning assignment.¹⁷ Some obvious candidates include blaming for ϕ -ing, disapproving of ϕ -ing, avoiding ϕ -ing, but there may also be others. With these in mind, in the next section, I will explore the issue of the logical possibility of moral dilemmas within Schroeder’s semantic framework.

3. How expressivists can and should secure the logical possibility of moral dilemmas

As I noted in Section 1, in this paper I tentatively assume that an expressivist semantic account should aim to secure the logical possibility of moral dilemmas—as a semantic theory, it should respect the data that ordinary speakers do not take the conjunction of “ ϕ -ing is obligatory” and “not ϕ -ing is obligatory” to be an utterly confused non-sensical statement. This at least requires that *obligation implies permission* should not turn out

¹⁶ Köhler (2017) also highlights the essentially structural nature of Schroeder’s proposal to defend it from an objection raised by Skorupski (2012).

¹⁷ There is a very weak material requirement on what kind of unary expression one can put into the placeholder: it should be, at least, some negative attitude/act toward the object of evaluation. (I thank an anonymous reviewer for pointing this out.) For example, putting *praising* into the placeholder would yield *being for praising for ϕ -ing*, which is structurally adequate but obviously implausible as a meaning assignment for “ ϕ -ing is wrong”. Notice that all of the candidates Schroeder considers (i.e., blaming for ϕ -ing, disapproving of ϕ -ing, avoiding ϕ -ing) meet this requirement.

to be formally valid in the theory (again, there are also other combinations of axioms that one would need to invalidate, which I will not discuss in this paper—see Section 1). In this section, I will explore how one can achieve this task within Schroeder’s framework.

The discussion will proceed in two steps. First, I will show that whether one can successfully achieve this task in Schroeder’s framework ultimately depends on which specific kind of attitude/act one decides to put in the placeholder in the attitude assigned the meaning of “ ϕ -ing is wrong”, i.e., *being for [one’s preferred unary expression] ϕ -ing*. Second, I will argue that it will be crucial for expressivists that their decision here should not turn out to be problematically *ad hoc*. Seen from a broader perspective, to decide which specific act/attitude should go into the placeholder is to answer a very basic question for expressivists: what is the non-cognitive attitude expressed by “ ϕ -ing is wrong”, after all? This is a question whose answer should be motivated by general semantic-psychological considerations, *not just* by whether or not the resulting theory can secure the logical possibility of moral dilemmas. The case study will eventually show that this in fact makes it difficult for expressivists to invalidate *obligation implies permission* to get the right technical result in their theory. My final goal will be to locate the ultimate source of the difficulty in the mentalist assumption of the expressivist semantic project itself.

3.1 The logical possibility of moral dilemmas in Schroeder’s expressivist semantics

So far, following Schroeder, I have been focusing on the predicate “is wrong”. To address the question whether one can make *obligation implies permission* formally invalid in Schroeder’s framework, it is necessary to translate all the sentences that contain “is wrong” to the sentences that contain “is obligatory/permisible”. The required translation is shown in the following table.

ϕ -ing is wrong	ϕ -ing is not permissible	Not ϕ -ing is obligatory
Not ϕ -ing is wrong	Not ϕ -ing is not permissible	ϕ -ing is obligatory
ϕ -ing is not wrong	ϕ -ing is permissible	Not ϕ -ing is not obligatory
Not ϕ -ing is not wrong	Not ϕ -ing is permissible	ϕ -ing is not obligatory

Table 1

The proof of the table only requires two uncontroversial assumptions: (a) “Not ϕ -ing is obligatory” is a translation of “ ϕ -ing is wrong”, and (b) obligation and permission are interdefinable (i.e., “ ϕ -ing is permissible”

can be defined as “Not ϕ -ing is not obligatory” and *vice versa*). These directly yield the result that “ ϕ -ing is not permissible” is a translation of “ ϕ -ing is wrong” (represented in the first row), and one can similarly prove the rest of the table.¹⁸

I believe that these are natural assumptions in extending Schroeder’s analysis of “is wrong” to “is obligatory/permissible”. In Schroeder’s framework, (a) amounts to the idea that “ ϕ -ing is wrong” and “Not ϕ -ing is obligatory” express the same being for attitude, namely, FOR(blaming for ϕ -ing). (b) amounts to the idea that “ ϕ -ing is permissible” and “Not ϕ -ing is not obligatory” express the same attitude, namely, FOR(not blaming for ϕ -ing). In fact, once one accepts (a), it is unclear how one can avoid (b) in Schroeder’s framework: to reject (b), one would have to maintain that the pairs of the sentences in each of the rows (such as “ ϕ -ing is permissible” and “Not ϕ -ing is not obligatory”) express non-equivalent being for attitudes. It is questionable whether one can come up with any reasonably simple assignment of being for attitudes that meets this condition. (And, as I noted in footnote 6, Schroeder himself takes the interdefinability of obligation and permission to be uncontroversial anyway.)

As one can see from the table, the claim that “ ϕ -ing is obligatory” implies “ ϕ -ing is permissible” translates to the claim that “Not ϕ -ing is wrong” implies “ ϕ -ing is not wrong”. As I discussed in Section 1, whether *obligation implies permission* turns out to be formally valid in one’s semantic theory is an important question, because of its connection to the classic puzzle concerning the logical possibility of moral dilemmas. The puzzle was that the logical possibility of genuine dilemmas conflicts with the apparently plausible theses, each of which one might be inclined to treat as an axiom in one’s semantic theory: (1) permission can be defined in terms of obligation, and (2) obligation implies permission. As noted above, (1) is an independently plausible assumption, and, specifically in Schroeder’s framework, it is difficult to find a way to reject it. Hence, one should either maintain that (2) *obligation implies permission* is not formally valid or admit that moral dilemmas are impossible as a matter of logic and semantics alone. As I will explain below, this also tracks how Schroeder himself pursues this matter.

At one point in the book, Schroeder (Ch. 5, Sec. 4) discusses a possible treatment of *obligation implies permission* in his framework, in connection

¹⁸ Here, I assume that the definition for negation (NEG) provided in the previous section is applicable to sentences with different predicates than “is wrong”. Another important point is that one need not use the axiom *obligation implies permission* to prove this table.

with the issue of the logical possibility of moral dilemmas. There, Schroeder simply registers the fact that there are theorists who believe that moral dilemmas should be logically *impossible*, and he goes on to explore whether his semantics could accommodate such a claim. Unlike me, Schroeder does not make any assumption concerning whether or not expressivists in general should aim to secure the logical possibility of moral dilemmas in their semantics. (In the next subsection, I will explain the ramifications of the divergence in stance here.) His aim is, rather, to show that his semantic framework is compatible with either of the opposing views on this issue: it “can remain neutral on this question [about the logical (im)possibility of moral dilemmas], offering ways for those who like either result to capture their views” (74-5).¹⁹ Specifically, Schroeder maintains that there is a way to “supplement our system with an auxiliary assumption that will yield the result that ‘murdering is wrong’ [i.e., ‘not murdering is obligatory’] and ‘not murdering is wrong’ [i.e., ‘murdering is obligatory’] turn out to be inconsistent” (72). So, on Schroeder’s view, one can either adopt or reject the “auxiliary assumption” in question to reflect one’s preferred view on the logical possibility of moral dilemmas. Below, let me introduce Schroeder’s “auxiliary assumption” and explain how it will make moral dilemmas logically impossible in the current framework.²⁰

The auxiliary assumption in question states that “blaming for not murdering entails not blaming for murdering” (73). According to Schroeder, this validates *obligation implies permission* in the current framework. Later, I will question exactly what the auxiliary assumption is claiming in substance and how one might justify it—for now, let us simply confirm the technical point first. Notice that (given Table 1) “ ϕ -ing is obligatory” and “ ϕ -ing is permissible” express the following attitudes, respectively.

ϕ -ing is obligatory (not ϕ -ing is wrong)	FOR(blaming for not ϕ -ing)
ϕ -ing is permissible (ϕ -ing is not wrong)	FOR(not blaming for ϕ -ing)

Table 2

The auxiliary assumption in question states that the following holds:

(AA) Blaming for not ϕ -ing entails not blaming for ϕ -ing.

¹⁹ Schroeder’s neutral stance toward this issue is also reflected in his comment that it should “pay to be cautious about building this [i.e., the result that ‘murdering is wrong’ and ‘not murdering is wrong’ turn out to be inconsistent] into our logic” (72, ft. 6).

²⁰ Unfortunately, the proof of this point is only sketched and is not fully worked out by Schroeder himself—I will try to remedy it here. I thank an anonymous reviewer for pointing out the need for making the proof more explicit.

This, Schroeder claims, results in the following entailment relation between being for attitudes that captures *obligation implies permission*:

(OP) FOR(blaming for not ϕ -ing) entails FOR(not blaming for ϕ -ing).

Strictly speaking, to move from (AA) to (OP), one would require an assumption that for any pair of being for attitudes, entailment relations that hold between the contents of the attitudes will reflect in the corresponding entailment relations between the attitudes (that is, FOR(α) entails FOR(β) if and only if α entails β). Schroeder does not seem to explicitly discuss this, but let us accept it as a generalization of Schroeder's basic proposal that being for attitudes are inconsistency transmitting-attitudes—they are, recall, attitudes that are inconsistent with one another if and only if their contents are inconsistent. This is motivated by the general idea that being for attitudes are, like beliefs, attitudes such that their logical relationships (such as inconsistency) are reducible to the logical relationships that hold between the embedded contents. If one can assume this much, then the auxiliary assumption in question does capture *obligation implies permission* in Schroeder's framework.

Now, let us confirm how this will rule out the possibility of moral dilemmas. A moral dilemma is, again, a situation where ϕ -ing and not ϕ -ing are both obligatory for an agent. Below, I will demonstrate that “ ϕ -ing is obligatory and not ϕ -ing is obligatory” and *obligation implies permission* are jointly inconsistent in Schroeder's framework. The dilemma's conjuncts, “ ϕ -ing is obligatory” and “not ϕ -ing is obligatory”, express the following attitudes, respectively:

“ ϕ -ing is obligatory” expresses FOR(blaming for not ϕ -ing).

“Not ϕ -ing is obligatory” expresses FOR(blaming for ϕ -ing).

Assigning the meaning for the dilemmatic statement, “ ϕ -ing is obligatory and not ϕ -ing is obligatory”, requires introducing the definition for conjunction that Schroeder provides on page 66:

(AND) If ‘A’ expresses FOR(α) and ‘B’ expresses FOR(β), ‘A&B’ expresses FOR($\alpha \wedge \beta$).

Accordingly, “ ϕ -ing is obligatory and not ϕ -ing is obligatory” expresses the following attitude:

(MD*) FOR(blaming for not ϕ -ing and blaming for ϕ -ing).

Now, it needs to be shown that (MD*) and (OP) are jointly inconsistent. To proceed from here, one only needs to assume that having the attitude of FOR($\alpha\wedge\beta$) is equivalent to having the attitudes of FOR(α) and FOR(β). If this can be assumed, the proof is obvious. Having the attitude (MD*) amounts to having the following pair of the attitudes, (1) FOR(blaming for not φ -ing) and (2) FOR(blaming for φ -ing). (1) and (OP) immediately yield FOR(not blaming for φ -ing). This is directly inconsistent with (2). Hence, (OP) and (MD*) are jointly inconsistent.

This should suffice to show Schroeder is right to claim that the auxiliary assumption in question rules out the logical possibility of moral dilemmas in his framework. According to Schroeder, one can then either adopt or drop the auxiliary assumption depending on one's view on the logical possibility of moral dilemmas. Now, the question I want to pursue below is this. Can an expressivist really adopt or drop the auxiliary assumption that flexibly, as Schroeder seems to assume? A potential worry stems from the point that I set aside earlier. In claiming that *blaming for not murdering entails not blaming for murdering*, one actually seems to be making a substantive claim about blaming. That is, whether it is true or not seems to depend on what blame actually is, or how the notion of blame should be understood. If so, the auxiliary assumption is making a claim whose plausibility may need to be examined *independently of* one's view on the logical possibility of moral dilemmas. This might mean that expressivists cannot in fact adopt or drop the auxiliary assumption as they like to deal with the logical possibility of moral dilemmas. (In the current paper, unlike Schroeder, I am assuming that expressivists generally need to side with the pro-dilemma view. So, the question I will focus on in the next subsection is this: can expressivists reject the auxiliary assumption freely, just because they need to invalidate *obligation implies permission* and make moral dilemmas logically possible?)

Schroeder, in fact, seems to recognize this sort of concern himself. In the same section, Schroeder points out that if one wishes to adopt the auxiliary assumption to make moral dilemmas logically impossible in the proposed semantic framework, one may need to justify it by maintaining that "it is [as a matter of conceptual necessity] impossible to both blame for murdering and blame for not murdering" (73). This, in effect, is to justify the auxiliary assumption by maintaining that it expresses a conceptual truth about blaming (as Schroeder puts it, a truth in the "logic of blaming" (73)). In passing, however, Schroeder also notes that this may actually seem "a little too strong for plausibility" (74). Although Schroeder does not elaborate on this, certainly we may imagine someone who insists that one can consistently blame someone for not φ -ing and blame the same person for φ -ing. For example, in Sophie's choice, Sophie is forced to choose only

one child from the two, and, whichever child she ends up choosing, she might blame herself for not choosing the other. To take a more mundane example, one can imagine, say, a poor heavy smoker who will be blamed by her family and friends anyway regardless of whether she continues smoking or refrains from doing so.²¹ The existence of this kind of practice concerning blame can certainly cast doubt on the idea that the auxiliary assumption expresses a conceptual truth about blame.

Generalizing from this point, I think one can arrive at an important observation: the plausibility of the auxiliary assumption in question depends on one's decision as to which act/attitude should go into the placeholder in the attitude of *being for [some unary expression] ϕ -ing*. Schroeder puts this point in this way:

A different idea about 'wrong' is that 'murder is wrong' expresses being for avoiding murdering. On this account, the assumption required to yield the inconsistency is that it is impossible to both avoid murdering and avoid not murdering, which *is*, in fact, a highly plausible assumption about the logic of avoiding. So how easy it is to get 'murdering is wrong' and 'not murdering is wrong' to turn out to be inconsistent will obviously turn on which account we give of the attitude expressed by 'murdering is wrong'. (74)

The auxiliary assumption originally states: "blaming for not murdering entails not blaming for murdering" (73). If one decides that *avoiding*, instead of *blaming*, should go into the placeholder, the auxiliary assumption would have a different content, correspondingly: *avoiding not murdering entails not avoiding murdering*. Here, the latter might actually appear more plausible than the former, because it is highly unintuitive to think that an agent can avoid not ϕ -ing and avoid ϕ -ing at the same time—this looks similar to the case of intending ϕ -ing and intending not ϕ -ing at the same time, which seems impossible or at least deeply irrational. This in turn means that it can be highly controversial to *reject* this version of the auxiliary assumption (i.e., *avoiding not ϕ -ing entails not avoiding ϕ -ing*), because it appears to capture an independently plausible claim that follows from the "logic" of avoiding. On the other hand, if one instead decides that, say, *disliking* should be put into the placeholder, one would get *disliking not ϕ -ing entails not disliking ϕ -ing* as the corresponding auxiliary

²¹ Imagine the following: if she smokes after dinner, her family might blame her for doing so by claiming that it harms their health; if she decides to refrain from smoking, her family might blame her for not smoking, claiming that there is an important value to sticking with one's habit and she should not be influenced so easily by others' advice. So, they will blame her anyway. This, at least, seems to capture what people do sometimes.

assumption. This may seem rather implausible—we can coherently imagine a universal hater who dislikes pretty much everything, including both φ -ing and not φ -ing. It would of course depend on how one understands the notion of disliking in question, but the point is that rejecting this version of the auxiliary assumption seems less controversial compared to the different version that one gets by putting *avoiding* in the placeholder.

Let me summarize the current point by connecting it to my exposition of Schroeder's semantics in the previous section. To adequately deal with the negation problem, one's expressivist semantic theory needs to meet the structural requirement that the attitude assigned as the meaning of "ϕ-ing is wrong" should take the following form: *being for [some unary expression] ϕ-ing*. Recall that this in itself does not call for any strong material restriction on which specific kind of expression should go into the placeholder. But now, there is at least one important consideration that one should take into account in making one's decision here: the plausibility of the auxiliary assumption, which validates *obligation implies permission*, depends on which specific attitude/act gets plugged into the placeholder. I think this is an interesting result that one can extract from Schroeder's discussion, which is worth pressing further than he actually does. Building on this point, in the next subsection I will explore how expressivists should ultimately deal with this issue and explain how this seemingly technical point actually exposes a more general problem for the expressivist semantic project.

3.2 Basic meaning assignment and its empirical implications

As I mentioned, Schroeder neither endorses nor rejects the auxiliary assumption himself—he merely presents it as an option that one can either adopt or reject, depending on one's view on the logical possibility of moral dilemmas. Although Schroeder's neutral stance is justified given the overall aim of his discussion, I think one can actually push this point further than Schroeder himself does to pose a general challenge for proponents of expressivism. One can do so by, for the sake of argument, sharing the assumption of the current paper that it is in fact a requirement for expressivists to secure the logical possibility of moral dilemmas. As I explained, this in turn requires (at least) invalidating *obligation implies permission* in one's semantic theory. So, although Schroeder simply allows expressivists to either accept or reject the auxiliary assumption, one can advance the discussion further by assuming that they are actually committed to rejecting it.

This leads to an important question. From the discussion above, we know that the plausibility of the auxiliary assumption depends on which specific attitude/act gets plugged into the placeholder in *being for [some unary expression] φ -ing*, assigned as the meaning of “ φ -ing is wrong (is not permissible)”. Now, can expressivists justifiably decide to put e.g., *blaming* instead of *avoiding* into the placeholder, *solely on the basis of the fact* that this would make it easier for them to reject the auxiliary assumption and thereby invalidate *obligation implies permission*? The answer I defend below is *no*. More specifically, I argue that their decision about which specific act/attitude must go into the placeholder should be criticized as problematically *ad hoc*, if it is motivated *only* by the need for securing the logical possibility of moral dilemmas.

To see why, notice first that the question of which act/attitude should go into the placeholder is a question that concerns the theory’s basic meaning/attitude assignment for an atomic sentence that contains its target normative expression: what is the non-cognitive attitude that ordinary speakers express by sincerely uttering, “ φ -ing is wrong”, after all? Here, recall also that expressivists are committed to the mentalist assumption about the meanings of normative sentences (see Section 2): the non-cognitive attitudes assigned as the meanings of normative sentences are *mental states* of speakers who express them via their utterances. Therefore, in deciding what to put into the placeholder in *being for [some unary expression] φ -ing*, expressivists are making a *substantive empirical-psychological claim* about the mental states that underlie the use of “is wrong” in the actual linguistic practice.²² As such, naturally, their theoretical decisions need to be empirically well-motivated. Therefore, if securing the logical possibility of moral dilemmas in their semantic theory is the *only* reason for their decision in their basic meaning assignment for “is wrong”, it is problematically *ad hoc*; this is because it simply ignores other equally important, notably psychological, considerations that expressivists need to take into account in motivating their basic meaning assignment.

Let me demonstrate this point in more concrete terms. Their decision in the basic meaning assignment will, for example, yield predictions concerning what kind of behavioral patterns are generally compatible with one’s sincere utterance of “ φ -ing is wrong”. The plausibility of their decision should be then tested by examining whether the predictions it yields fit with ordinary speakers’ actual behaviors as well as their intuitions on this matter. Suppose that one’s expressivist semantic theory tells us that, as Schroeder supposes, a speaker’s sincere utterance of “ φ -ing is wrong”

²² I thank an anonymous reviewer for urging me to make this point more explicit.

expresses the attitude of *being for blaming for φ -ing*. This proposal has an implication that a speaker who sincerely utters this sentence must be generally disposed to sanction actual instances of φ -ing performed by others. Surely, one who possess the attitude of being for blaming for φ -ing must feel compelled to actually blame others' performances of φ -ing, at least when circumstances permit. Likewise, people should be likely to find puzzling a situation where a speaker sincerely utters " φ -ing is wrong" but never cares to blame observed instances of φ -ing at all. The question is: does this in fact capture ordinary speakers' behavior and their intuitions on this matter? If yes, putting *blaming* into the placeholder is empirically well-motivated—if no, the choice may need to be reconsidered.

Now, contrast this to an alternative proposal, which assigns the attitude of FOR(*avoiding φ -ing*) as the meaning of " φ -ing is wrong".²³ This account now yields a different prediction concerning how ordinary speakers would react to the kind of situation described above. This is because avoiding φ -ing may be, unlike blaming, a matter of making personal plans for oneself, which may not necessarily concern whether one would also publicly sanction *others'* performances of φ -ing. If this is so, my having the attitude of FOR(*avoiding meat-eating*), for example, might simply mean my being committed to avoid eating meat myself (and, perhaps, vaguely hope others do the same). This attitude, unlike FOR(*blaming for meat-eating*), need not imply that I am committed to socially sanction those who do not act as I do. Accordingly, even if I am known for overtly asserting "Meat consumption is wrong", my not taking any corrective actions toward those who continue to consume meat need not appear so puzzling on the alternative proposal. Again, the question is: does this actually fit with ordinary speakers' behavior and intuitions?

This quick comparison between the two choices above should suffice to illustrate how one's decision about which attitude/act should be put into the placeholder needs to be motivated by general empirical-psychological considerations concerning the use of "is wrong" in ordinary speakers' linguistic practice. Whatever decisions they end up making in their basic meaning assignment for "is wrong", the mentalist assumption of the expressivist semantic project implies that they are *also* making claims about individual speakers' psychology. As such, they yield various predictions that need to be tested empirically in light of the data. What I discussed above is just one example, and I suspect that there are also other

²³ As I explained in the previous section, choosing to assign FOR(*avoiding φ -ing*), instead of FOR(*blaming for φ -ing*), as the meaning of " φ -ing is wrong" leads to the result that the auxiliary assumption *obligation implies permission* will appear more plausible. The current point is that this kind of choice, when taken together with the mentalist assumption, also yields other predictions that should not be ignored.

types of similar considerations that one should take into account (e.g., to what extent do ordinary speakers take a speaker's sincere utterance of "φ-ing is wrong" to be compatible with her emotional neutral reactions to instances of φ-ing?). It is for this reason that, in deciding which specific act/attitude should go into the placeholder in *being for [some unary expression] φ-ing*, expressivists cannot simply insist that they are justified to choose whatever act/attitude that would invalidate *obligation implies permission* and make moral dilemmas logically possible.

It might be helpful to elaborate on the current point by connecting it to the fact that when Schroeder chooses to specifically assign FOR(blaming for φ-ing) as the meaning of "φ-ing is wrong", he purports to be following the proposal by Gibbard (Schroeder 2008a, 58). For Gibbard, "to call something rational is to express one's acceptance of norms that permit it" (Gibbard 1990, 7); and, accordingly, "φ-ing is irrational" (which should be read as "φ-ing is wrong", given his overall picture) would express a state of *accepting a norm that forbids φ-ing*, which looks similar to *being for blaming for φ-ing*.²⁴ Again, Schroeder is not necessarily committed to this specific choice, and he draws on Gibbard just to "fix examples" (58). However, surely Gibbard himself should have some basic reasons and motivations (including considerations such as above) for analyzing "is wrong" ultimately in terms of blame/forbiddance, instead of avoidance, disapproval, disliking and so on.²⁵ And this means that Gibbard (in his 1990 book) and others who adopt the notion of blame/forbiddance in analyzing the meaning of "is wrong" are *prima facie* committed to accepting whatever theoretical consequences that follow from "the logic of blame/forbiddance". If it tells them that blaming for (forbidding) not φ-ing and blaming for (forbidding) φ-ing are inconsistent, they are *prima facie* committed to accepting its consequence in their semantic theory: *obligation implies permission* turns out to be formally valid, which in turn makes moral dilemmas logically impossible. My contention is that even if something like this turns out to be the case, they cannot easily switch to a different analysis that invokes e.g., the notion of avoidance instead of blame/forbiddance just for the purpose of blocking this result. Such a response should be criticized as problematically *ad hoc*. As I argued, securing the possibility of moral dilemmas is only one of the

²⁴ More precisely, for Gibbard, moral judgements are "judgments of what moral feelings it is rational to have", that is, "judgments of when guilt and resentment are apt" (1990, 6). Gibbard then analyzes an act of calling something rational or irrational in terms of a speaker's expression of acceptance (which is a non-cognitive mental state) of norms that permit or forbid the object of evaluation. So, Gibbard's analysis of "is wrong" is expressivist in somewhat indirect way, mediated by his expressivist understanding of the evaluation of rationality. I believe, however, that the overall plausibility of the discussion does not depend on the details of Gibbard's theory.

²⁵ One might take issue with this point—perhaps, Gibbard may have no deep reason to invoke the notion of forbiddance in his analysis of "is irrational". I will address this point at the end of this section.

considerations that one should take into account in determining the basic meaning assignment for “is wrong” in one’s theory. If one hastily makes changes in the basic attitude assignment to deal with this particular technical problem, it is likely to produce unintended predictions in other places and even runs the risk of unintentionally abandoning whatever basic insights that motivated one’s theory in the first place.

The current point can be generalized. Different expressivist accounts invoke different basic notions in analyzing the meaning of “is wrong” (or whatever normative predicate or operator that they take to be basic, such as “is irrational”, “ought”, “is obligatory”, etc.). To take a few examples, Blackburn (1984, 1988) analyzes “ ϕ -ing is wrong/impermissible” in terms of *booing/disapproving ϕ -ing*; Gibbard (2003) analyzes “ ϕ -ing is the thing to do” in terms of a state of *planning to ϕ* , Horgan and Timmons (2006) analyze “One ought to ϕ ” in terms of *an ought-commitment that one ϕ ’s*. Each of these different proposals should be motivated by some basic theoretical considerations that they take to be important, including observations of ordinary speakers’ behavior and intuitions concerning the use of the target expressions. Depending on which of these proposals one finds plausible and which act/attitude one thinks should be put in the placeholder in *being for [some unary expression] ϕ -ing*, different results will follow as to whether moral dilemmas are logically possible. Even if they do not like the result, modifying their basic attitude assignment just for the purpose of blocking it would be problematically *ad hoc*. I have demonstrated this point in some detail, focusing on an expressivist semantic account that invokes the notion of blaming in its basic meaning assignment. I believe that one can pose, *mutatis mutandis*, the same point for any kind of expressivist semantic theory.²⁶

²⁶ Let me briefly demonstrate this point focusing on Blackburn’s proposal as an example. Blackburn stipulates, following Ayer, that the meaning of “ ϕ -ing is wrong/impermissible” is a speaker’s disapproval of ϕ -ing (Blackburn 1984, 195). Suppose that one is now convinced that the notion of disapproval should be invoked in one’s expressivist semantic analysis of “is wrong”. Since Blackburn’s original proposal faces the negation problem, one would need to reformulate Blackburn’s proposal using Schroeder’s framework—one obvious way to do so is to think that “ ϕ -ing is wrong” expresses FOR(disapproving ϕ -ing). Here, if one thinks that there are good reasons to believe that it is not inconsistent to disapprove ϕ -ing and disapprove not ϕ -ing at the same time, one would have to accept that FOR(disapproving not ϕ -ing) *does not* entail FOR(not disapproving ϕ -ing). As a result, *obligation implies permission* turns out to be invalid in this Blackburn-inspired semantics, and, accordingly, moral dilemmas turn out to be logically possible (again, assuming that the theory does not validate other combinations of axioms that make moral dilemmas logically impossible). If expressivists should side with the pro-dilemma view, the result must be a welcoming one for those who find Blackburn’s choice generally convincing. Of course, if one thinks that disapproving ϕ -ing and disapproving not ϕ -ing *are* inconsistent, then the opposite result will follow. That is exactly my point—whichever turns out to be the case, one cannot simply change the basic meaning assignment for “ ϕ -ing is wrong/impermissible” *just because* one wants to avoid some particular result. Any such move should be criticized as problematically *ad hoc*.

Before closing this section, let me address one potential objection. In the discussion above, I assumed that expressivists such as Blackburn and Gibbard have some independent theoretical reasons and motivations to stick with specific notions (such as disapproval, blame/forbiddance and so on) in their analyses of “is wrong”. One might find this assumption dubious and object that their choices are not really based on any substantive, let alone empirical, considerations, because their primary aim is merely to construct *structurally adequate* expressivist semantic accounts that can deal with the Frege-Geach problem (although, if Schroeder is correct, they do not succeed in achieving this aim either). For example, Blackburn does not provide any lengthy discussion to justify his choice—he merely notes in passing that he is following Ayer (Blackburn 1984, 167). So, one might say, expressivists are free to switch to whatever attitude/act that seems suitable for dealing with technical problems at hand (such as making moral dilemmas logically possible) and there is nothing *ad hoc* about this move.

My response to the objection would be that expressivists including Blackburn, Gibbard and others *should have* supported their choices by some non-trivial empirical-psychological considerations, even if they in fact did not do so. As I explained, when taken together with the mentalist assumption of the expressivist semantic project, one’s choice in the basic meaning assignment will yield various predictions that should be empirically tested, whether they like it or not. I demonstrated this point by comparing analyses that invoke different notions (such as blame, avoidance) in their basic meaning assignments. The discussion in this section, if successful, shows that expressivists cannot remain indifferent to this issue and simply maintain that whatever attitude/act will do as long as it allows them to deal with technical problems in their semantic theories. It has to be recognized that, in the expressivist semantic project, one’s decision in the basic meaning assignment always comes with psychological implications.

4. Concluding remarks: The mentalist assumption of expressivism and its costs

In this paper, I explored how expressivists can secure the logical possibility of moral dilemmas in their semantic theories, using Schroeder’s framework as a case study. Even if one’s expressivist semantic theory is structurally adequate in that it can deal with the Frege-Geach problem, securing the logical possibility of moral dilemmas remains as a separate task. Specifically, in Schroeder’s framework, whether or not moral dilemmas turn out to be logically possible depends on which specific attitude/act one thinks should go into the placeholder in the attitude of

being for [some unary expression] ϕ -ing, assigned as the meaning of “ ϕ -ing is wrong”. Due to the mentalist assumption of the expressivist semantic project, deciding what should be put into the placeholder involves making a substantive empirical claim about the psychology that underlies the use of “is wrong” in the actual linguistic practice. Expressivists then need to take many things into consideration in making their decision, including, for example, actual behavioral patterns that typically follow a speaker’s sincere utterance of “ ϕ -ing is wrong” and folk intuitions on this matter. Accordingly, their decision should be criticized as problematically *ad hoc* if it is solely motivated by the need for invalidating *obligation implies permission* to secure the logical possibility of moral dilemmas.

Let me conclude by articulating the moral of the case study in more general terms. Overall, the case study suggests that the difficulty for expressivists mainly derives from the mentalist assumption of their semantic project that logico-semantic relations exhibited by normative sentences should be captured in terms of the psychological attitudes that speakers express by uttering them. Whenever expressivists wish to make a certain theoretical move to deal with a technical problem (e.g., invalidating *obligation implies permissibility* to make moral dilemmas logically possible), they first need to confirm that their move is consistent with the basic meaning assignment in their semantic theory. If adopting the desired theoretical move requires changing the basic meaning assignment, expressivists will need to commit to whatever empirical-psychological claims entailed by such a change. This seems to capture how the mentalist assumption generally prevents expressivists from flexibly adopting or dropping axioms in their theories to get the right technical results.

Recall also that, as I noted in Section 1, there are other combinations of intuitively plausible axioms that are known to be inconsistent with the logical possibility of moral dilemmas (e.g., *obligation implies possibility* and the principle of agglomeration).²⁷ Extending the strategy of the current paper, one can similarly examine whether one’s preferred expressivist semantic theory can find reasonable ways to avoid such combinations. The discussion in this paper might give the impression that the prospect is indeed dim.

Of course, there is a more general question: which aspects of the data should one’s semantic theory aim to respect in the end? After all, the logical possibility of moral dilemmas and *obligation implies permission* are both intuitively plausible, and ordinary speakers may often behave as

²⁷ I thank an anonymous reviewer for suggesting to explore the implications of the case study from this angle.

if both are true. Their actual linguistic practice exhibits inconsistencies in some places, and a formal semantic theory, if it purports to be consistent, will have to ignore some aspects of the data. What needs to be given up must be decided based on many considerations, and it could turn out that, all things considered, it is better to give up the logical possibility of moral dilemmas instead of *obligation implies permission* in one's semantic theory. This is a problem for every semanticist, not just for expressivists. The moral of the paper is that expressivists need to face an *extra* constraint in addressing this kind of issue: the mentalist assumption of their semantic project prevents them from flexibly dropping or adopting axioms in their theory to deal with technical problems. This, I believe, provides an explanation of why expressivists *in particular* will have hard time addressing technical issues such as the treatment of the logical possibility of moral dilemmas. And, importantly, this point would hold even if it turns out that expressivists are not required to make moral dilemmas possible in their theories, as I assumed. Regardless of whether they ultimately need to validate or invalidate *obligation implies permission* in their theories, the crucial point is that whatever theoretical moves necessary for arriving at the desired result will need to be justifiable in light of the mentalist assumption. The case study in this paper has shown that this is a significant burden that expressivists need to bear in pursuing their semantic project.²⁸

Acknowledgments

I am grateful to Paul Bloomfield and two anonymous reviewers for the journal for their valuable comments on earlier versions of this paper.

Funding information: This work was supported by the Japan Society for the Promotion of Science (JSPS KAKENHI 21J00685 and 22KJ0522).

REFERENCES

- Ayer, Alfred J. 1936. *Language, Truth, and Logic*. Dover.
- Almeida, Michael J. 1990. "Deontic Logic and the Possibility of Moral Conflict." *Erkenntnis* 33 (1): 57–71.
<https://doi.org/10.1007/BF00634551>.
- Baker, Derek, and Jack Woods. 2015. "How Expressivists Can and Should Explain Inconsistency." *Ethics* 125 (2): 391–424.

²⁸ Does this mean that any expressivist semantic account is bound to collapse at some point due to its mentalist assumption? If one thinks so, then the result of the current paper might be taken as providing indirect support for attempts at exploring alternative, *non*-semantic ways of cashing out basic expressivist ideas (see Bar-On and Chrisman 2009; Bar-On et al. 2014).

- <https://doi.org/10.1086/678371>.
- Bar-On, Dorit, and Matthew Chrisman. 2009. "Ethical Neo-Expressivism." *Oxford Studies in Metaethics* 4: 133–66.
- Bar-On, Dorit, Matthew Chrisman, and James Sias. 2014. "(How) Is Ethical Neo-Expressivism a Hybrid View?" In *Having It Both Ways: Hybrid Theories and Modern Metaethics*, edited by Guy Fletcher and Michael Ridge, 223–47. Oxford University Press.
- Bar-On, Dorit, and James Sias. 2013. "Varieties of Expressivism." *Philosophy Compass* 8 (8): 699–713.
<https://doi.org/10.1111/phc3.12051>.
- Blackburn, Simon. 1984. *Spreading the Word: Groundings in the Philosophy of Language*. Clarendon Press.
- . 1988. "Attitudes and Contents." *Ethics* 98 (3): 501–17.
<https://doi.org/10.1086/292968>.
- . 1998. *Ruling Passions: A Theory of Practical Reasoning*. Oxford University Press.
- Brink, David O. 1994. "Moral Conflict and Its Structure." *The Philosophical Review* 103 (2): 215.
<https://doi.org/10.2307/2185737>.
- Conee, Earl. 1982. "Against Moral Dilemmas." *Philosophical Review* 91 (1): 87–97. <https://doi.org/10.2307/2184670>.
- Geach, Peter T. 1960. "Ascriptivism." *Philosophical Review* 69 (2): 221–25. <https://doi.org/10.2307/2183506>.
- . 1965. "Assertion." *Philosophical Review* 74 (4): 449–65.
<https://doi.org/10.2307/2183123>.
- Gibbard, Allan. 1990. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Harvard University Press.
- . 2003. *Thinking How to Live*. Harvard University Press.
- Goble, Lou. 2009. "Normative Conflicts and The Logic of 'Ought.'" *Noûs* 43 (3): 450–89. <https://doi.org/10.1111/j.1468-0068.2009.00714.x>.
- Hansson, Sven Ove. 2019. "In Defence of Deontic Diversity." *Journal of Logic and Computation* 29 (3): 349–67.
<https://doi.org/10.1093/logcom/exv057>.
- Hare, R. M. 1952. *The Language of Morals*. Oxford University Press.
- . 1970. "Meaning and Speech Acts." *Philosophical Review* 79 (1): 3–24. <https://doi.org/10.2307/2184066>.
- Holbo, John. 2002. "Moral Dilemmas and the Logic of Obligation." *American Philosophical Quarterly* 39 (3): 259–74.
- Horgan, Terry, and Mark Timmons. 2006. "Cognitivist Expressivism." In *Metaethics after Moore*, edited by Terry Horgan and Mark Timmons, 255–98. Oxford University Press.
- Köhler, Sebastian. 2017. "The Frege-Geach Objection to Expressivism, Structurally Answered." *Journal of Ethics and Social Philosophy* 6 (2): 1–7. <https://doi.org/10.26556/jesp.v6i2.150>.

- Lemmon, E. J. 1962. "Moral Dilemmas." *Philosophical Review* 71 (2): 139–58. <https://doi.org/10.2307/2182983>.
- Marcus, Ruth Barcan. 1980. "Moral Dilemmas and Consistency." *The Journal of Philosophy* 77 (3): 121–36. <https://doi.org/10.2307/2025665>.
- McConnell, Terrance C. 1978. "Moral Dilemmas and Consistency in Ethics." *Canadian Journal of Philosophy* 8 (2): 269–87.
- . 2022. "Moral Dilemmas." In *The Stanford Encyclopedia of Philosophy* (Fall 2022 Edition), edited by Edward N. Zalta. Retrieved February 22, 2024. <https://plato.stanford.edu/archives/fall2022/entries/moral-dilemmas/>
- McNamara, Paul, and Frederik Van De Putte. 2023. "Deontic Logic." In *The Stanford Encyclopedia of Philosophy* (Fall 2023 Edition), edited by Edward N. Zalta. Retrieved February 22, 2024. <https://plato.stanford.edu/archives/fall2023/entries/logic-deontic/>
- Nair, Shyam. 2016. "Conflicting Reasons, Unconflicting 'Ought's.'" *Philosophical Studies* 173 (3): 629–63. <https://doi.org/10.1007/s11098-015-0511-4>.
- Ross, W. D. 1930. *The Right and the Good*. Oxford University Press.
- Schroeder, Mark. 2008a. *Being for: Evaluating the Semantic Program of Expressivism*. Clarendon Press.
- . 2008b. "What Is the Frege-Geach Problem?" *Philosophy Compass* 3 (4): 703–20. <https://doi.org/10.1111/j.1747-9991.2008.00155.x>.
- . 2008c. "How Expressivists Can and Should Solve Their Problem with Negation." *Noûs* 42 (4): 573–99. <https://doi.org/10.1111/j.1468-0068.2008.00693.x>.
- . 2010. *Noncognitivism in Ethics*. Routledge.
- Sias, James. 2024. "Ethical Expressivism." *The Internet Encyclopedia of Philosophy*, edited by James Fieser and Bradley Dowden. Retrieved March 13, 2024. <https://iep.utm.edu/eth-expr/>
- Sinnott-Armstrong, Walter. 1988. *Moral Dilemmas*. Blackwell.
- Skorupski, J. 2012. "The Frege-Geach Objection to Expressivism: Still Unanswered." *Analysis* 72 (1): 9–18. <https://doi.org/10.1093/analys/anr136>.
- Stevenson, Charles L. 1937. "The Emotive Meaning of Ethical Terms." *Mind* 46 (181):14–31.
- Styron, William. 1979. *Sophie's Choice*. Random House.
- Tessman, Lisa. 2015. *Moral Failure: On the Impossible Demands of Morality*. Oxford University Press.
- Unwin, Nicholas. 2001. "Norms and Negation: A Problem for Gibbard's Logic." *The Philosophical Quarterly* 51 (202): 60–75.

- Williams, Bernard. 1966. "Consistency and Realism." *Proceedings of the Aristotelian Society (Supplement)* 40: 1–22.
- Vallentyne, Peter. 1989. "Two Types of Moral Dilemmas." *Erkenntnis* 30 (3): 301–18. <https://doi.org/10.1007/BF00168283>.
- Van Fraassen, Bas C. 1973. "Values and the Heart's Command." *Journal of Philosophy* 70 (1): 5–19. <https://doi.org/10.2307/2024762>.
- Zimmerman, Michael J. 1996. *The Concept of Moral Obligation*. Cambridge University Press.

