

A Frege-Geach Style Objection to Cognitivist Judgment Internalism

Thorsten SANDER[†]

This is the peer reviewed version of the following article:

T. Sander: A Frege-Geach Style Objection to Cognitivist Judgment Internalism, *Dialectica* 68: 391-408, which has been published in final form at <<https://doi.org/10.1111/1746-8361.12075>>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

ABSTRACT

According to (cognitivist) judgment internalism, there is a conceptual connection between moral judgment and motivation. This paper offers an argument against that kind of internalism that does not involve counterexamples of the amoralist sort. Instead, it is argued that these forms of judgment internalism fall prey to a Frege-Geach type argument.

1. Introduction

Traditional motivational internalism is the view that there is an "internal" (i.e. conceptual) relation between moral judgment and motivation: it is part of the *meaning* of moral statements or part of the *concept* of a moral judgment that a person who sincerely judges that an action is right will have a corresponding motive. Here are two well-known statements of this claim:

Internalism is the view that ... the motivation must be so tied to the truth, or meaning, of ethical statements that when in a particular case someone is ... morally required to do something, it follows that he has a motivation for doing it. (Nagel 1970, 7)

An internalist theory is a theory according to which the knowledge ... of a moral judgment implies the existence of a motive (not necessarily overriding) for acting on that judgment. If I judge that some action is right, it is implied that I have, and acknowledge, some motive or reason for performing that action. It is part of the sense of the judgment that a motive is present: if someone agrees that an action is right, but cannot see any motive or reason for

[†] Universität Duisburg-Essen, Fakultät für Geisteswissenschaften, Institut für Philosophie, 45117 Essen. Germany, Email: thorsten.sander@uni-due.de.

doing it, we must suppose, according to these views, that she does not quite know what she means when she agrees that the action is right. (Korsgaard 1986, 8-9)¹

However, the general claim that the acceptance "of a moral judgment implies the existence of a motive ... for acting on that judgment" (Korsgaard 1986, 8-9) seems to be much too strong. Consider the following sentences:

- (1) Richard Nixon is evil.
- (2) I should have helped her.

In a natural sense of the expression "moral judgment" uttering one of these sentences clearly amounts to making such a judgment, but under normal circumstances a judgment made by uttering (1) seems to be motivationally inert, and judgments like (2) are at best loosely linked to the utterer's motivation. In both cases there is no connection between the propositional content of the judgment and the content of the utterer's motive (if there is any).

In order to deal with this problem, motivational internalism has to be restricted to a certain class of moral judgments, i.e. to those judgments whose propositional contents deal with an action of the utterer in the context of the utterance (here and now or in the not too distant future). This idea leads us to the following characterisation of motivational internalism:

(MI) [It is a conceptual truth that:] If an agent judges that it is right for her to ϕ in circumstances C, then she is motivated to ϕ in C. (Smith 1994, 61)²

The kind of internalism that is expressed by (MI) has been frequently contested by offering actual or possible counterexamples. Putative cases involve persons who are wicked, who do not care about moral considerations ("amoralists"), who suffer from listlessness or persons with brain damage.³ If these critics are right, (MI) is either not a conceptual truth or not even true at all. In the following I will, however, pursue a different strategy. Instead of offering yet

¹ There are, of course, numerous non-cognitivist who also subscribe to internalism, but non-cognitivist internalism is not my subject here.

² The phrase in square brackets is my addition. Instead of the term "conceptual truth", Smith (in his 1994 book) speaks of a "conceptual connection" between judgment and motivation; in a later paper (2008, 211), however, Smith does use the term "conceptual truth" in his characterisation of (weak) motivational internalism and calls internalism an "a priori constraint on what is to count as a moral judgment" (2008, 210).

³ See, respectively, Stocker (1979); Brink (1989), 45-50; Mele (1996); Roskies (2003).

another counterexample to (MI), I will rather try to show that those kinds of motivational internalism that, in some way or another, embody a semantic claim, i.e. a claim about the *sense* or the *meaning* of terms such as "right" or "ought" are at odds with the semantics of moral discourse. More specifically: they fall prey to a Frege-Geach (FG) style objection. (When, in what follows, I speak of *internalism*, *judgment internalism* or *motivational internalism*, the terms are to be understood as applying only to those forms of internalism that actually advance such a semantic or conceptual claim. In order to prevent misunderstandings, I will sometimes use the more specific term *semantic internalism*.)

2. The Argument

I shall start with a somewhat simplified version of my argument, in application to the somewhat simple (and strong) form of internalism that is captured by the principle (MI). (The argument in its full form will emerge in the next two sections.) Consider a completion of the sentence form "It is right for me to ϕ ":

(3) It is right for me to fight in this war.

According to semantic internalism, the motivational force is, in some sense, part of the very meaning of (3). When somebody utters (3) sincerely and in accordance with the sentence's literal meaning, she thereby expresses having a motive to fight. But what about a sentence of the form "If it is right for me to ϕ , then it is right for me to χ "? Suppose somebody is determined to fight in a war against a "rogue state" and, in thinking about that issue, reasons thusly:

(P1) It is right for me to fight in this war.

(P2) If it is right for me to fight in this war, then it is right for me to kill human beings.

(C) So, it is right for me to kill human beings.

While it is quite plausible to suppose that a speaker who sincerely utters the sentence (P1) cannot fail to be motivated (at least to a certain degree), a speaker who makes the conditional judgment (P2) clearly does not have to have a motive for taking part in a war. Uttering a conditional sentence about the relation between the rightness of two different actions does not imply that one is motivated to perform the action mentioned in the antecedent of the

conditional. Thus, if the motivational force is part of the meaning of (P1), it seems as though the sentence "It is right for me to fight in this war" cannot have the same meaning in (P2), where it is embedded, and in (P1), where it appears free-standing. Accordingly, a person who argues in the above-mentioned, seemingly innocuous way would have to be accused of committing a fallacy of equivocation.

So there seems to be a serious objection against motivational internalism that does not depend on any (disputable) assumptions about the moral psychology of actual or possible human beings. We do not even have to rely on some sophisticated semantic theory; we only need Frege's theory of predication to show that internalism is based on a problematic account of the semantics of moral discourse, i.e. on an account that smuggles an essentially pragmatic or force-related element of moral utterances into their meaning. Although statements of the form "It is right for me to ϕ ", "I ought to ϕ " etc. are typically used to *acknowledge* our duty to ϕ , to *express* our intention or our having a motive to do so, only the additional claim that this way of using moral utterances can serve as a key to understanding the semantics of moral statements will lead us to the conclusion that the motivational force is in some sense part of the *meaning* of such sentences.⁴

One can readily grant that a sentence form such as

(4) It is right for me to ϕ , but I am not motivated to ϕ

is, in some sense, odd. But the oddity of such an utterance does not have to be explained by the assumption that, as internalism would have it, a speaker who utters (4) commits herself to a literal contradiction, just as a speaker who utters a sentence of the form

(5) I know that p , but I do not believe that p .

Instead of that, one can explain the oddity of (4) in much the same way as some philosophers (see Martinich 1980, 224-225) would account for the absurdity of

(6) p , but I do not believe that p ,

⁴ Kalderon has coined the term "pragmatic fallacy" for the non-cognitivist's (alleged) mistake to confuse "the contents of moral sentences with what their utterances normally convey" (2005, 65). I am not sure whether non-cognitivism is actually based on such a fallacy, but if it is, then semantic internalism seems to commit the pragmatic fallacy as well.

i.e. by appealing to the Gricean concept of a conversational implicature.⁵ By uttering "It is raining" one conversationally implies that one believes that this is the case (although this is clearly not part of the meaning of that sentence), and by uttering "It is right for me to ϕ " one conversationally implies that one is, to some degree, motivated to perform the action (without it being the case that the motivational force adheres to the meaning of the sentence).⁶

Motivational internalism thus seems to inherit one of the basic flaws of analytic philosophy as it was practised around the middle of the last century: an unconvincing equation of meaning and use that fails to properly distinguish between semantics and pragmatics, combined with a one-sided diet that focuses on only one way of using expressions like "right", "ought" etc., i.e. only in declarative sentences in the first person singular present tense.

3. Four Objections

Everybody familiar with the current debate on metaethics will have noticed that the above argument against motivational internalism is, in some sense, not at all new: a structurally similar argument is known as the Frege-Geach objection to non-cognitivism (see Geach, 1958, 54; Geach, 1960; Geach, 1965; Searle, 1962). Now it is a truism that there is a close connection between internalism and non-cognitivism. Not only do the most prominent advocates of non-cognitivism in fact subscribe to internalism, but they also regard the idea that a certain degree of motivation is "built into" moral judgments as one of the best arguments for non-cognitivism. There are, however, philosophers like Thomas Nagel, John McDowell, Christine Korsgaard, Michael Smith or Ralph Wedgwood who support certain versions of internalism, while eschewing any commitment to non-cognitivism. So one may wonder whether one of the standard objections to non-cognitivism can also serve as an argument against internalism in general. Aren't there fundamental differences between these

⁵ One might argue that the oddity of (4) is, rather, due to *conventional* implicature. Since the whole distinction between conventional and conversational implicatures is highly contested (see, e.g., Bach, 1999), I cannot offer a full defence of my claim, but (granted that cancellability is the most reliable criterion for conversational implicatures) there seems to be a simple argument in favour of the view that motivational force is *not* conventional: there are cases in which the motivational upshot of a moral judgment *can* be cancelled without generating a contradiction, i.e. those cases in which a speaker offers a reason for her not being motivated. The sentence "I (really) ought to ϕ , but I am so depressed that I cannot motivate myself to do anything" would be an example. Utterances of the form (4), in contrast, seem to betray a simple lack of interest in normative concerns.

⁶ See Finlay (2004) who claims that in many contexts speakers who make moral judgments conversationally implicate that they have a motive; see also Bar-On and Chrisman (2009) for a somewhat similar approach according to which motivational force is a matter of expression "in the action sense".

two positions that make it impossible to use the FG point as it has been used here? Let me discuss four possible objections in order to clarify the scope of the argument.

Objection 1: Philosophers like Simon Blackburn (1984, 189-96), Alan Gibbard (1990, 83-102) and Mark Schroeder (2008a) have offered sustained defences of non-cognitivism against the FG objection. Why should it be impossible to meet the FG style objection to internalism in much the same way? Furthermore, recent years have seen the development of diverse "hybrid" or "ecumenical" theories in metaethics (see, e.g., Ridge, 2006 and Boisvert, 2008) which seem to promise a "cheap" solution to the FG problem in general. And, finally, one might question or modify the very distinction that underlies the Frege point, i.e. Frege's distinction between thought and illocutionary force (see Dummett 1981, 328-354; Hom and Schwartz, 2013; Sander, forthcoming).

Response: It is not my aim in this paper to evaluate the cogency of the FG objection or the success of the different proposals to defuse the objection. My claim is, rather, of a conditional nature: if the FG point constitutes the foundation of a successful argument against non-cognitivism, then we can use the very same point to demonstrate that cognitivist internalism is in trouble, too. So, I do not wish to say that the objection *cannot be met* by cognitivist internalists, only that, in the face of the FG objection, cognitivist internalism *does not fare much better* than non-cognitivism. Since none of the aforementioned strategies to deal with the FG argument are without problems⁷, meeting the challenge will require some serious philosophical work.

Objection 2: Non-cognitivists claim that moral judgments are *synonymous* to some non-assertive (directive or expressive) speech act that *lacks a truth value*. Internalism, on the other hand, is neither committed to any claim about the synonymy of moral sentences with other kinds of sentences nor to the claim that moral statements do not have a truth-value. Isn't the FG objection just an argument against *performative* theories according to which the meaning of sentences (involving predicates like "right", "true" etc.) can be characterised by their potential to be used in performing illocutionary acts of a certain kind?

⁷ The difficulties in constructing a "logic of attitudes" have received much attention (for an overview, see Miller 2003, 58-73; 95-107). Schroeder (2009) offers a critical survey of the "virtues and vices" of hybrid theories.

Response: It is abundantly clear that the non-cognitivist thesis that "Murder is wrong" means the same as, say, "Do not murder!" raises some serious difficulties with respect to inferences such as:

(P3) Murder is wrong.

(P4) If murder is wrong, then promoting murder is wrong.

(C2) Thus, promoting murder is wrong.

First, substituting (P3) with the allegedly synonymous sentence "Do not murder!" yields a sequence of sentences that, just for *syntactic* reasons, does not seem to be an inference at all; imperatives cannot function as the antecedent of a conditional, and we do not know how to deduce a conclusion from an imperative and a normal conditional statement (see Geach 1958, 53). Secondly, if moral judgments, as non-cognitivism would have it, lack a truth-value and if, as is widely held, only sentences that are true or false can serve as premises or conclusions of deductive inferences, there is an additional (logico-)*semantic* reason why the non-cognitivist is not able to account for the apparent validity of the above inference.⁸

Although these two points are fairly convincing arguments to the effect that non-cognitivism is at odds with our practice of using moral judgments as premises or conclusions of inferences, these arguments are only loosely connected to the original FG point. The FG objection and the two arguments above (that indeed work only against an opponent who subscribes to a claim about synonymy or lack of truth-value) both see the major flaw of non-cognitivism in its inability to properly account for the inferential use of moral statements. However, in order to be dialectically successful, the FG argument does *not* require an opponent who actually claims that moral sentences lack a truth value as their meaning is *exhausted* by their emotive or prescriptive force. For the argument to work, it is sufficient that the opponent claims that the emotive or prescriptive force *is part of* the meaning of some moral utterances.⁹ This force would then be a semantic property of moral utterances that is cancelled in embedded contexts, which leads to the familiar problem that, on the non-cognitivist or internalist account, expressions like "good" or "right" cannot have the same meaning when embedded as when unembedded, and that, in turn, raises the challenge of how

⁸ That imperatives are not truth-apt is one of the standard objections to imperative logic (see Jørgensen, 1937), and the same criticism can be levelled at non-cognitivism.

⁹ Searle (1962, 424) explicitly directs his argument to an opponent who claims that it is "part of the meaning [of an expression] that any speaker who utters a sentence containing it ... is characteristically performing act A".

to account for the apparently correct and successful use of moral judgments as part of inferences.

Objection 3: In the above argument it is claimed that, according to internalism, a speaker who utters the sentence "It is right for me to ϕ " thereby *expresses* having a motive to ϕ . This characterisation of internalism already presupposes that internalism is a kind of expressivism. And so no wonder that, on this reading, internalism falls prey to the FG point. But internalism is, in fact, not committed to the view that in making moral judgments we *express* a motivational state; internalism just claims that a speaker who makes a moral judgment necessarily *has* a motive to ϕ . Being motivated is, unlike an emotion or a desire, not a mental state that a person expresses by making moral judgments.

Response: Although I do think it would be apt to call internalism a kind of "(hybrid) motive expressivism"¹⁰, the argument does not essentially depend on that assumption, and so, fortunately, I do not have to delve into the deep questions that are connected to the two terms "motive" and "expression". Whatever a motive is, and whatever it means to have or to express a motive, the internalist claims that "having" a motive is part of the very meaning of moral utterances (of a certain kind), and if we consider contexts in which moral sentences appear as constituents, e.g. as part of a conditional, this inevitably leads to the kind of challenge raised by the FG point.

Objection 4: Transposing Geach's original argument against ascriptivism, the performative theory of truth, non-cognitivism etc., into one against internalism works only if internalism is seen as a theory of the meaning of certain *sentences*. But isn't internalism rather a theory of the motivational upshot of our moral *judgments*?¹¹

Response: It is quite evident that internalism cannot be a thesis *only* about the meaning of certain sentences. To see why, consider an actor who utters a sentence of the form "It is right for me to ϕ " on stage. If we assume, rather uncontroversially, that the Fregean sense of the sentence does not change in such a scenario, there seems to be only one plausible way of

¹⁰ For characterisations of internalism that invite such a reading see Frankena (1976, 60); Korsgaard (1986, 8-9).

¹¹ There are quite a few authors who use the term "judgment" in their characterisation of motivational internalism (Frankena 1976, 49; Korsgaard 1986, 8; Nagel 1970, 7; Smith 1994, 61) or who speak of "judgment internalism" instead of "motivational internalism" (Darwall 1985, 54; Wedgwood 2007, 25-27).

explaining why the actor does not commit herself to anything: the illocutionary force of an assertion is missing. The actor utters certain words without making the judgment one would ordinarily make by uttering the words. Thus, in order to get a viable version of internalism, the motivational upshot must be a function of the meaning of sentences that contain words like "right" or "ought" *and* of the fact that these are uttered with a specific illocutionary force. And that is, I assume, what Christine Korsgaard (1986, 9) means when she says that it "is part of the sense of the judgment that a motive is present".

So, at first glance, there seems to be a clear contrast between internalism and the theories Geach actually attacks. While these theories embody a claim about the meaning of certain words or sentences, internalism is a thesis about judgments with a specific propositional content. But now recall Geach's "distinction between calling a thing 'P' and predicating 'P' of a thing" (Geach 1960, 223). We can, for example, use the predicate "true" without calling something "true" and, in the same way, we can use the predicate "right" without actually calling something "right" (and without undertaking any further commitments).

The adherents of those theories the Frege-Geach argument is directed at do present their theories as an analysis of the meaning of words like "true", "good" etc., but the very point of Geach's argument is that this way of presenting things is utterly mistaken. Performative theories do not (or do not successfully) analyse the *meanings* of these words; at best, they say something correct about the "*assertoric content*" (Dummett 2004, 32) of sentences that contain these words. But, in actual fact, internalism is also a theory of the assertoric content of certain sentences that contain predicates like "right", and internalists do not say anything about the question of how this kind of content is related to the meaning of "right" in other contexts (most notably, in the contexts Geach discusses). Thus, the invocation of judgments is not a way of evading the FG problem for internalism. Rather, the claim that only judgments with a specific propositional content have a motivational upshot is exactly what leads to the problem.

4. Weak Motivational Internalism

So far, I have been discussing only a comparatively strong version of internalism according to which *all* judgments of the form "It is right for me to ϕ " or "I ought to ϕ " necessarily involve some degree of motivation. This kind of internalism, however, seems to have gone largely out of fashion. In order to evade the difficulties raised by externalist counterexamples,

contemporary internalists often resort to a more restricted claim. One such version of internalism is presented by Smith (1994, 61):¹²

(MI*) [It is a conceptual truth that:] If an agent judges that it is right for her to ϕ in circumstances C, then either she is motivated to ϕ in C or she is practically irrational.

Since (MI*) does not claim that a moral judgment *on its own* carries motivational force, the principle does not seem to be susceptible to the FG argument. This impression, however, is mistaken. To see why, recall our modus ponens argument from section 2:

(P1) It is right for me to fight in this war.

(P2) If it is right for me to fight in this war, then it is right for me to kill human beings.

(C) So, it is right for me to kill human beings.

While it is true that, according to (MI*), sincere utterances of (P1) do *not* have to carry motivational force, the problem remains essentially the same. If (MI*) is conceptually true, it is part of the meaning of (P1) that somebody who sincerely utters this sentence is either motivated to fight *or practically irrational* (or, alternatively, that a *rational* speaker who utters (P1) is motivated to fight), and it should be clear that the embedded occurrence of (P1), as the antecedent of (P2), does not have that kind of meaning. Thus, given that the semantic properties of (P1) and of (P1) as part of (P2) seem to be different, the proponent of (MI*) has to explain how the "embedded meaning" of "It is right for me to ϕ " is related to the "free-standing meaning" of the very same sentence. A proponent of (MI*), or of a similar principle, cannot just *stipulate* that the meanings of the terms "judging", "right (for)", "motive" etc. compose in such a way that (MI*) comes out as (conceptually) true.

It seems evident that this problem is in no way dependent on the *specific kind* of property of the agent which might sever the link between judgment and motivation. Thus, the result can be generalised to all versions of internalism that conform to the following pattern:

(MI**) It is a conceptual truth that: If an agent judges that it is right for her to ϕ , then either she is motivated to ϕ or she is F,

¹² For Smith's defence of this "practicality requirement" see Smith (1994, 71-76).

and so the differences between various forms of weak or conditional internalism are irrelevant to the purpose of this paper.¹³

Thus, the challenge cognitivist internalism faces is not really different from the traditional FG challenge to non-cognitivism. In order to solve the FG problem, the non-cognitivist has to

... provide a compositional semantics ... which tells us the meaning of complex sentences of a certain kind in terms of the meanings of the parts of that sentence ... and ... to show that this is an adequate semantics for the sentence, because it can predict and explain the sentence's semantic properties. (Schroeder 2008, 708)

In order to meet the challenge that has been presented here, cognitivist internalism has to do essentially the same thing.

In a next step, I will present a reason why it will be rather difficult for the internalist to cope with this challenge: the semantic property that is essential to the internalist's point of view, i.e. motivational force, seems to be so vague and context-dependent that internalism would need a *highly* sophisticated semantic theory in order to "predict and explain" that semantic property. Let me start with a comparison: one of the standard complaints against "epistemic theories of meaning" such as verificationism or inferentialism is that these theories cannot properly account for the compositional structure of natural languages. If, so the argument goes, meaning is something epistemic like a sentence's assertibility conditions and if meanings are compositional while assertibility conditions are not, then meanings cannot be identified with something epistemic (see Fodor and Lepore 1992, 175-186). For a simple example, consider the two sentences

(7) I shall give you a D.

(8) I intend to give you a D.

If (semantic) verificationism is the claim that sameness of verification conditions implies sameness of meaning, verificationism seems to be committed to the thesis that (7) and (8) have the same meaning: whatever counts as a reason for (7) will also be a reason for (8) and *vice versa*. But this cannot be the whole story, for, if we embed (7) and (8) into certain

¹³ For a helpful survey see Björklund et. al. (2012, 126-128).

conditionals, the resulting sentences seem to have different meanings (see Dummett 2004, 33 for this example):

(9) If I give you a D, you will forfeit your grant.

(10) If I intend to give you a D, you will forfeit your grant.

This problem certainly does not amount to a knock-down argument against verificationism; rather, it should be seen as a challenge to that position. However, in order to solve this problem, it is not enough just to distinguish two "levels" of meaning by saying that (7) and (8) have the same "assertoric content", but not the same "ingredient sense" (for these terms, see Dummett 2004, 33), for that is nothing more than a restatement of the problem. In order to actually answer the challenge, it is necessary for the verificationist to explain how exactly assertoric content and ingredient sense are related. What are the conditions under which two sentences have the same ingredient sense? How does the ingredient sense of (7) or (8) determine the assertoric content of more complex sentences like (9) or (10)? In the absence of any satisfying answers to these questions, verificationism is, at best, a nice idea, but not a fully-fledged theory of meaning.

But it is not only epistemic theories of meaning that face such a challenge. Right at the beginning of this paper, it was noted that it would be utterly implausible to suppose that *all* moral judgments are internally related to the utterer's motivation; at best, internalism is valid for a restricted class of moral judgments that, arguably, includes sentences like "It is right for me to ϕ ", but not "I should have helped her". This problem gets even more serious if we bear in mind that there are *numerous ways* of depriving the sentence "It is right for me to ϕ " of its motivational force: changing the pronoun, using a different tense ("It was right for me to ϕ "), modifying the illocutionary force ("Is it right for me to ϕ ?"), and – the point that leads to the FG problem for internalism – embedding the sentence in a conditional. The internalist then owes us an account of why only certain sentences involving moral terms like "right" carry motivational force and of how the moral terms contribute to the meaning of the sentences of which they are constituents. Basically, the problem is this: not all occurrences of "It is right for me to ϕ " (or the like) carry motivational force, while there are moral sentences that do *not* involve a first-person pronoun and that nevertheless seem to carry motivational force. Can the

internalist devise a compositional semantic theory that predicts and explains which moral sentences, exactly, have this semantic property?¹⁴

There are, admittedly, linguistic contexts in which such problems do not even seem to arise. If a (rational) thinker makes a judgment by uttering a sentence of the form "If it is right for me to ϕ , then it is right for me to χ ", she does not have to be motivated at all; and if a (rational) speaker utters the sentence "It is right for me to ϕ , and it is right for me to χ ", she must have a motive to ϕ and a motive to χ (granted that a principle like (MI*) is true). So it seems as though logical connectives determine the motivational upshot of complex moral judgments in the same way that they determine the "assertoric upshot" of complex assertions. (While asserting " p and q " amounts to asserting p and asserting q , asserting "if p , then q " does not mean asserting p or asserting q .)

But there are other contexts in which things are not that simple. Let us take disjunctions like "It is right for me to ϕ , or it is right for me to χ " as a first example. The analogy between assertions and moral judgments would suggest that even a rational thinker's judgment with such content does not commit her to being motivated at all. (Asserting " p or q " is not equivalent to asserting p and asserting q .) But is this really the case? Can one actually judge that it is right to ϕ *or* to χ without having a motive (not necessarily overriding) to ϕ *and* to χ ? Or are disjunctive moral judgments motivationally inert? The internalist owes us an account of how disjunctive moral judgments are related to the judger's motives.

A second phenomenon that creates difficulties is tense. I already noted that judgments like "I should have helped her" do not seem to be internally related to the judger's motives. But what about judgments of the form "It will be right for me to ϕ (at some time t)"? Does that mean that the judger is already motivated to ϕ and is just waiting for the time where her motive can eventually be put into action? This hardly seems credible, since, in a case where t is a point of time in the distant future, this would imply that one can have a "dormant" motive for a long period of time. However, claiming that the future tense cancels any motivational

¹⁴ It seems as though even hybrid theories that do not try to offer a *meaning analysis* of moral judgments are susceptible to this kind of problem. Bar-On's and Chrisman's "neo-expressivism" (2009) is based on the idea of enforcing "a separation between the semantic content of a claim considered as a product and the mental state that gets expressed by the person who makes the claim" (2009, 150): while moral *sentences* "s-express truth-evaluable propositions" (159), *speakers* "a-express a motivational state" (143) by uttering such sentences. Bar-On and Chrisman then try to vindicate a certain form of internalism: "what is distinctive about ethical claims... is the fact that a person who issues an ethical claim is supposed to give voice to a ... motivational state using a linguistic ... vehicle that involves ethical terms". (144) Since, clearly, not all sentences that involve ethical terms carry motivational force even in this sense (sentence (1) in the first section of this paper would be an example), neo-expressivists have to explain which linguistic vehicles are such that uttering such a sentence without having a motivational state would mean "violating one of the *propriety conditions* on making ethical claims" (143).

upshot is not an attractive option either, for then we could easily devise a sorites-style argument. Suppose t_1 is a point of time in the distant future and t_0 designates the time of utterance. If a rational agent who utters the sentence "It will be right for me to ϕ at t_1 " at t_0 does not have to be motivated to ϕ and if a very small period of time (say, one millisecond) cannot make a difference with respect to the utterer's motivation, then we should assume that "It will be right for me to ϕ in one second" does not carry any motivational force either. But, as far as motivation is concerned, there does not seem to be any substantial difference between "It will be right for me to ϕ in one second" and "It is right for me to ϕ ". Presumably, these difficulties are the chief reason why, in Michael Smith's (1994, 61) characterisation of internalism, "circumstances" are taken into account. If, as in (MI*), the judgment and the motive have to occur in the same circumstances, moral judgments about our future actions should be motivationally inert.¹⁵ As "circumstance" is a highly vague term, this claim, however, leads directly to the above-mentioned sorites problem. How far, we may ask, does the judger's motive have to reach beyond the context of the judgment?

Finally, recall the piece of practical reasoning which, as I claimed at the beginning of this paper, leads to the FG problem for internalism. Assume now that (P1), (P2) and (C) are the thoughts of a rational speaker. Then, given (MI*), it should be uncontroversial that an utterance of "It is right for me to fight in this war" carries motivational force while uttering "If it is right for me to fight in this war, then it is right for me to kill human beings" or its antecedent does not. But what about the conclusion "So, it is right for me to kill human beings"? Does somebody who makes that judgment as a conclusion of the inference have to be motivated to kill human beings? If there were reason to believe that motivational force is, as it were, closed under deduction, the judger would have to have that motive, but such a "motive closure principle" seems to be even more questionable than its epistemic counterpart. Isn't it possible that the judger simply does not have a corresponding motive or that, on realising the consequences of her moral judgments, actually loses any motivation to fight in the war? (See Shafer-Landau 2000, 273 for a somewhat similar example.) Once again, the internalist owes us an account: of how and whether the motivational force of judgments is preserved from the premises to the conclusion.

As in the case of verificationism, these considerations, of course, do not amount to a refutation of internalism, but they do present an important challenge to that position. A

¹⁵ Strictly speaking, (MI*) is not directly applicable to such cases, since there is no place to insert an auxiliary verb indicating the future tense. But I do think that this way of putting things clearly captures the "spirit" of Smith's proposal.

satisfactory account of the relation between judgment, motivation (and rationality) would have to have an answer to all of the questions posed above. Since just knowing the principle (MI*), or similar principles, will not be of much help in performing that task, even the highly restricted version of motivational internalism Smith advocates has a problem with (the motivational upshot of) embedded moral judgments and is therefore in need of a genuine compositional semantics of moral terms.¹⁶

5. Wedgwood's "Normative Judgment Internalism"

So far, I have focused on two versions of internalism that are discussed in Michael Smith's *The Moral Problem*: a strong version Smith dismisses, and a weak version that is defended by him. This may lead to the suspicion that the problems I have presented are not really a problem for (semantic) internalism in general, but are due to some idiosyncratic elements in Smith's approach to that issue.

In order to defuse this suspicion, I shall finally discuss a more recent kind of internalism ("normative judgment internalism") which can be found in Ralph Wedgwood's *The Nature of Normativity*. With respect to the *semantics* of normative thought, Wedgwood's main idea is to provide an analysis of the concept *ought* by specifying the "concept's role in our thinking" (Wedgwood 2007, 78). As a matter of fact, Wedgwood's conceptual role semantics for the term "ought" is quite surveyable; if he is right, we should need only one rule in order to "characterize the essential conceptual role of the practical 'ought'" (Wedgwood 2007, 97):

(NJI*) Acceptance of the first-person proposition ' $O_{\langle me, t \rangle}(p)$ ' – where ' t ' refers to some time in the present or near future – commits one to making p part of one's ideal plan about what to do at t . (Wedgwood 2007, 97)

At first glance, there seems to be a vast difference between Smith's "practicality requirement" (MI*) and Wedgwood's rule (NJI*). On closer inspection, however, it can be seen that what distinguishes (NJI*) from (MI*) is just supposed to solve some minor (or major) "technical

¹⁶ Even though only those forms of internalism that assume a conceptual connection between judgment and motivation are directly vulnerable to the aforementioned type of problem, non-semantic internalists may be confronted with a broadly similar problem. Trivially, non-semantic or a posteriori internalism does not face the challenge of devising a compositional *semantics* of moral terms, but any such theory needs an answer to the question: which moral sentences, exactly, are empirically correlated with motivational states?

problems".¹⁷ Wedgwood's *intuitive* version of internalism, which would fall prey to these problems, can be found earlier in his book:

(NJI) Necessarily, if one is rational, then, if one judges 'I ought to ϕ ', one also intends to ϕ . (Wedgwood 2007, 25)

And there he also says that (NJI) is "close" to the versions of internalism that have been defended by Michael Smith and Christine Korsgaard. Thus, it should not come as a surprise that Wedgwood faces much the same challenge as Smith. To see why, consider a moral modus ponens argument of the following form:

- (P1) $O_{\langle me, t1 \rangle} (p)$
- (P2) If $O_{\langle me, t1 \rangle} (p)$, then $O_{\langle me, t2 \rangle} (q)$
- (C) $O_{\langle me, t2 \rangle} (q)$

While it is plausible that a thinker who accepts (P1) has to have a corresponding intention or is committed to making p part of her ideal plan, it should be clear that (NJI*) does not capture the meaning of "ought" when it appears as part of a more complex proposition. Uttering or thinking " $O_{\langle me, t1 \rangle} (p)$ " *as part of* (P2) does not appear to involve any commitment of that kind. Thus, Wedgwood's (2007, 97) claim that (NJI*) captures the "conceptual role of the practical 'ought'" seems to be false. (NJI*) indeed captures, to use Dummett's words, the assertoric content of *some* (i.e. the non-embedded) first-person propositions involving the concept *ought*; however, it does not tell us very much about the conceptual role of *ought* in general.¹⁸

Wedgwood seems to be aware of this problem; he even admits that (NJI*), in some sense, does *not* capture *all* aspects of the conceptual role of "ought":

¹⁷ These problems Wedgwood tries to solve by talking of "ideal plans" instead of "intentions" need not concern us here; see Wedgwood (2007, 29-31; 95-97).

¹⁸ Here is how Wedgwood (2007, 43) himself presents the FG problem for *expressivism*: "An adequate account of the meaning of normative terms must explain how they can figure, without a shift of meaning, both in statements in which they have largest scope, and embedded in subsentences of complex utterances in which they do not have largest scope." The problem is that Wedgwood's conceptual role semantics for the normative term "ought" is flawed in just the same way: (NJI*) specifies the meaning of "ought" only for those cases in which "ought" has largest scope.

The essential conceptual role of the practical 'ought' will have to entail NJI in the special case in which the proposition that is embedded in its scope is the proposition 'I ϕ ' ..., but it will have to be a more general conceptual role that includes other ways in which we are rationally required to use this concept. (Wedgwood 2007, 95-96)

However, Wedgwood does not give us any hint as to what such a general rule for "ought" would look like, and in striking contrast to his claim that there should be a rule that captures "a more general conceptual role" of "ought", he even maintains that the nature of that concept "is entirely determined by the rule" (NJI*) (Wedgwood 2007, 105). If we are meant to take this at face value, then there could not be any other rule that would (partially) determine the meaning of "ought", and this, in turn, would imply that the FG point does not only pose a challenge to Wedgwood's internalism; it would amount to a genuine refutation. For then the occurrence of " $O_{\langle me, t \rangle}(p)$ " in (P2) would lack any meaning whatsoever since the semantic content of "ought" in the embedded proposition is clearly not covered by the rule (NJI*).

If we leave aside Wedgwood's claim that the concept *ought* is "entirely determined" by (NJI*), how could he solve the FG style problem? There seem to be two options. Wedgwood could either devise a *general* rule that captures the conceptual role of the practical "ought" and that actually *entails* (NJI*). It is, however, far from obvious what such a general rule would look like. Alternatively, Wedgwood could stick with (NJI*), which captures the meaning of the genuinely practical "ought", and supplement it with other rules that take care of the cases in which "ought" does not seem to have a practical significance in the narrow sense. However, as long as (NJI*) forms part of the system of rules for the term "ought", it seems as though the resulting meaning theory will not be compositional.¹⁹ This worry is not just based on the general point that (NJI*) belongs to the species of conceptual role semantics and that conceptual role semantics has severe problems with compositionality (see, once again, Fodor and Lepore 1992, 175-186). Rather, Wedgwood's rule (NJI*) makes especially clear *why* certain versions of inferentialism suffer from this defect. Compare his rule (NJI*) to what is often considered to be the prime example of a comparatively successful form of inferentialism, i.e. to the inferentialist's explanation of the meaning of logical operators.²⁰ Such an explanation will comprise a set of rules that has to meet certain constraints: the rules should be "primitively compelling" (Peacocke 1992, 6), "self-justifying" (Dummett 1991, 251), "harmonious" (Dummett 1991, 215-220) and as independent of one another as

¹⁹ Wedgwood repeatedly stresses the compositionality constraint. See Wedgwood (2007, 86; 93).

²⁰ See Gentzen (1934-5); Peacocke (1992, 5-7); Brandom (2000, 61-63); Wedgwood (2007, 82-88).

possible.²¹ The result will be a neat system of natural deduction which provides one introduction and elimination rule for each operator and which, above all, clearly satisfies the compositionality principle.

(NJI*), however, does not seem to meet the compositionality constraint, since it is a rule that characterises a whole array of different concepts at once. Thus, it would serve as an elimination rule for the terms "accepting", "ought", "me", "present", "or", "near future" as well as an introduction rule for "committing", "part" and "ideal plan".²² (NJI*) does not explain how the meaning of, say, "I ought to ϕ " (and the commitments that are allegedly part of the meaning of this sentence) is determined by the subsentential expressions that are its constituents; (NJI*) rather characterises the assertoric content of those moral judgments that function as an internalist "intuition pump" as a whole and, thus, in a non-compositional way.²³

In the absence of any elaborate proposal for supplementing (NJI*) with other rules, I do not wish to maintain that such an approach would *inevitably* lead to a meaning theory that violates the compositionality constraint. Wedgwood, however, owes us a compositional meaning theory for the term "ought"; more specifically, he owes us a semantic theory that explains how exactly motivational force varies with different occurrences of "ought". If motivational force varies in a rather unpredictable manner, as I claimed in the last section, it will not be easy to devise such a meaning theory.

6. Conclusion

How far can we generalise this? The forms of internalism I have been discussing all seem to conform to a principle which might be summarised as follows:

(IP) It is a conceptual truth that: if an agent accepts that she stands in some normative relation to p , then [either] she will be in a motivational state with respect to p [or she is F].

²¹ To give just one example of the last constraint: though $\sim(p \ \& \ \sim q) \rightarrow \sim(q \rightarrow \sim p)$ is logically equivalent to p , it would not be a good idea to add a corresponding rule to a calculus, for then this rule would simultaneously determine the meaning of three different connectives.

²² In that respect, (NJI*) closely resembles Brandom's "material inferences" such as "I am a bank employee going to work, so I shall wear a necktie" (Brandom 2000, 84).

²³ Here is another way of stating the point. In a paper from 2010, Wedgwood explicitly describes his theory as neo-Fregean; more specifically, the Fregean sense of an expression is to be identified with the expression's conceptual role (2010, 480). However, given Frege's clear-cut distinction between sense and force, it seems evident that (NJI*) does not *just* capture the sense of moral sentences but also their (assertive) force.

Here, the term "normative relation" is a placeholder for different normative concepts like *ought* or *right*, and "motivational state" is a placeholder for concepts like *being motivated*, *intending* or *planning*, while anything that might sever the link between judgment and motivation (irrationality, or the like) is covered by *F*. Now compare (IP) to its "demodalised" counterpart:

(IP*) If an agent accepts that she stands in some normative relation to *p*, then either she will be in a motivational state with respect to *p* or she is *F*.

In contrast to (IP), (IP*) clearly does not raise any of the semantic problems I have focused on. Thus, the proponent of (IP*) only gets into serious trouble if she maintains that (IP*) captures an essential part of the *sense* of normative terms, for then she bears the burden of explaining how the sense of some normative term as it is captured by (IP) relates to other uses of such a term.

If Michael Smith (2008, 210) is right in his claim that the idea of there being a "conceptually necessary connection between moral judgment and moral motivation" is the "mark of internalism" (full stop), then the argument I have been developing would be an argument against internalism (full stop). Merely disputing over words, however, is of no profit, so let me note that I do not wish to claim that my argument amounts to a problem for internalism, broadly understood as a theory according to which there is *some* connection between moral judgment and motivation. Several theories that now go under the name of "internalism" are clearly not susceptible to the argument I have presented: internalism as an *a posteriori* claim, i.e. (IP*) or something similar, would be a clear example (see, e.g., Prinz 2006); another example would be "communal internalism" according to which, roughly, a failure to be motivated by one's moral judgment is possible, but "parasitic" on a communal practice in which agents are motivated by and act on their moral beliefs.²⁴

I am, however, pretty confident that *most* theories that regard internalism as a (conditional or unconditional) conceptual or *semantic* claim about the connection between individual judgment and *individual* motivation have to tackle the problems that I have discussed.*

²⁴ See Tresan (2009); see also Dreier (1990) for a somewhat similar approach.

* I thank Oliver Hallich, Miguel Hoeltje and Thomas Spitzley for their remarks on earlier drafts of this paper. I also thank five anonymous referees for *dialectica* whose comments helped me to clarify many points.

REFERENCES

- BACH, K. 1999, "The Myth of Conventional Implicature", *Linguistics and Philosophy* 22, pp. 327-366
- BAR-ON, D. and CHRISMAN, M. 2009, "Ethical Neo-Expressivism", in: Russ Shafer-Landau, ed., *Oxford Studies in Metaethics*, vol. 4, Oxford: Oxford University Press, pp. 133-165
- BLACKBURN, S. 1984, *Spreading the Word*, Oxford: Clarendon Press
- BOISVERT, D. 2008, "Expressive-Assertivism", *Pacific Philosophical Quarterly* 89, pp. 169-203
- BJÖRKLUND, F. et al. 2012, "Recent Work on Motivational Internalism", *Analysis Reviews* 72, pp. 124-137
- BRANDOM, R. 2000, *Articulating Reasons*. Cambridge, MA: Harvard University Press
- BRINK, D.O. 1989, *Moral Realism and the Foundations of Ethics*, Cambridge: Cambridge University Press
- DARWALL, S.L. 1985, *Impartial Reason*, Ithaca: Cornell University Press
- DREIER, J. 1990, "Internalism and Speaker Relativism", *Ethics* 101, pp. 6-26
- DUMMETT, M. 1981, *Frege. Philosophy of Language*, 2nd Edition, Cambridge: Cambridge University Press
- DUMMETT, M. 1991, *The Logical Basis of Metaphysics*, Cambridge, MA: Harvard University Press
- DUMMETT, M. 2004, *Truth and the Past*, New York: Columbia University Press
- FINLAY, S. 2004, "The Conversational Practicality of Value Judgement", *The Journal of Ethics* 8, pp. 205-223
- FODOR, J.A. and LEPORÉ, E. 1992, *Holism: a Shopper's Guide*, Oxford: Blackwell
- FRANKENA, W.K. 1976, "Obligation and Motivation in Recent Moral Philosophy", in: W.K. Frankena, *Perspectives on Morality*, Notre Dame: University of Notre Dame Press, pp. 49-73, 223-6
- GEACH, P.T. 1958, "Imperative and Deontic Logic", *Analysis* 18, pp. 49-56
- GEACH, P.T. 1960, "Ascriptivism", *The Philosophical Review* 69, pp. 221-225
- GEACH, P.T. 1965, "Assertion", *The Philosophical Review* 74, pp. 449-465
- GENTZEN, G. 1934-5, "Untersuchungen über das logische Schließen", *Mathematische Zeitschrift* 39, pp. 176-210; 405-431
- GIBBARD, A. 1990, *Wise Choices, Apt Feelings*, Oxford: Clarendon Press
- HOM, C. and SCHWARTZ, J. 2013, "Unity and the Frege-Geach Problem", *Philosophical Studies* 163, pp. 15-24
- JÖRGENSEN, J. 1937, "Imperatives and Logic", *Erkenntnis* 7, pp. 288-296
- KALDERON, M.E. 2005, *Moral Fictionalism*, Oxford: Clarendon Press
- KORSGAARD, C. 1986, "Skepticism about Practical Reason", *The Journal of Philosophy* 83, pp. 5-25
- MELE, A.R. 1996, "Internalist Moral Cognitivism and Listlessness", *Ethics* 106, pp. 727-753
- MARTINICH, A.P. 1980, "Conversational Maxims and some Philosophical Problems", *The Philosophical Quarterly* 30, pp. 215-228
- MILLER, A. 2003, *An Introduction to Contemporary Metaethics*, Cambridge: Polity Press
- NAGEL, T. 1970, *The Possibility of Altruism*, Princeton: Princeton University Press
- PEACOCKE, C. 1992, *A Study of Concepts*, Cambridge: Cambridge University Press
- PRINZ, J. 2006, "The Emotional Basis of Moral Judgments", *Philosophical Explorations* 9, pp. 29-43
- ROSKIES, A. 2003, "Are Ethical Judgments intrinsically Motivational? Lessons from 'Acquired Sociopathy'", *Philosophical Psychology* 16, pp. 51-66
- RIDGE, M. 2006, "Ecumenical Expressivism: Finessing Frege", *Ethics* 116, pp. 302-366
- SANDER, T. forthcoming, "Force Cancellation, Subscription and the Frege Point"

- SEARLE, J.R. 1962, "Meaning and Speech Acts", *The Philosophical Review* 71, pp. 423-432
- SCHROEDER, M. 2008, "What is the Frege-Geach Problem?", *Philosophy Compass* 3/4, pp. 703-720
- SCHROEDER, M. 2008a, *Being For: Evaluating the Semantic Program of Expressivism*, Oxford: Clarendon Press
- SCHROEDER, M. 2009, "Hybrid Expressivism: Virtues and Vices", *Ethics* 119, pp. 257-309
- SHAFFER-LANDAU, R. 2000, "A Defense of Motivational Externalism", *Philosophical Studies* 97, pp. 267-291
- SMITH, M. 1994, *The Moral Problem*. Oxford: Blackwell
- SMITH, M. 2008, "The Truth about Internalism", in: W. Sinnott-Armstrong, ed., *Moral Psychology*, vol. 3, Cambridge, MA: The MIT Press, pp. 207-215
- STOCKER, M. 1979, "Desiring the Bad: An Essay in Moral Psychology", *The Journal of Philosophy* 76, pp. 738-753
- TRESAN, J. 2009, "The Challenge of Communal Internalism", *The Journal of Value Inquiry* 43, pp. 179-199
- WEDGWOOD, R. 2007, *The Nature of Normativity*, Oxford: Clarendon Press