SOME REMARKS ON THE FREGE-GEACH EMBEDDING PROBLEM

by

Richard Anderson

B.A., University of Calgary, 2005

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS
in the Department
of
Philosophy

© Richard Anderson 2009
SIMON FRASER UNIVERSITY
Spring 2009

APPROVAL

Name: Richard Anderson

Degree: Master of Arts

Title of Thesis: Some Remarks on the Frege-Geach Embedding Problem

Examining Committee: Dr. Martin Hahn
Chair

_____

Dr. Francis Jeffry Pelletier
Professor, Philosophy (Emeritus)
Simon Fraser University
Senior Supervisor

_____

Dr. David Zimmerman
Professor, Philosophy
Simon Fraser University
Supervisor

_____

Dr. John Woods
Professor, Philosophy
University of British Columbia
External Examiner

Date Approved: _____March 30, 2009_____

Abstract

Expressivist theories of moral discourse deny that moral judgments express truth-apt propositions or that they correspond to moral facts. Rather, moral judgments are taken to express non-truth-apt and action-guiding attitudes of approval or disapproval. As a result, the classical accounts of validity, consistency and logical consequence cannot be directly applied to moral discourse. These logical limitations are exploited by the Frege-Geach embedding problem, which challenges expressivism to account for the fact that moral sentences can be embedded into truth-functional contexts, and that they can figure as premises in valid arguments. This thesis examines the embedding problem in detail, and analyzes two prominent expressivist responses to it: Simon Blackburn's logic of attitudes, and Allan Gibbard's normative logic. It will be argued that neither response presents a complete solution to the embedding problem. Then some alternative proposals will be investigated.

Keywords: Frege-Geach; Embedding problem; Meta-ethical expressivism; Normative logic.

Subject terms: Frege, Gottlob. 1848-1925; Language and languages – Philosophy.

Acknowledgments

Contents

Chapter 1

Expressivism and Embedding

Expressivist theories of moral discourse deny that moral judgments express truth-apt propositions that represent or describe states of affairs in the natural world. Rather, moral judgments are taken to express non-truth-apt attitudes of approval or disapproval: non-cognitive attitudes like desires, intentions, or motivational and action-guiding mental states. As a result, expressivism resists an effective analysis in standard truth-conditional semantics and truth-functional logic, as the classical accounts of truth, validity, consistency and logical consequence do not apply to non-truth-apt attitudes.

These logical and semantic limitations are exploited by the Frege-Geach embedding problem, which is generally taken to be a decisive refutation of expressivism. In its most general form, the problem challenges expressivism to square its restrictive logical and semantic commitments with the fact that moral sentences appear to function, at least on the surface, as if they have all of the semantic, logical and metaphysical properties of ordinary descriptive sentences.

This chapter examines meta-ethical expressivism and the Frege-Geach embedding

problem. The structure of the chapter is as follows: in §1.1, the distinction between moral and descriptive discourse will be sharpened by attributing to the expressivist position two basic theses, Nonfactualism and the Expression Thesis, which are taken to characterize the semantic and logical commitments of expressivism. In §1.2, the Frege-Geach embedding problem will be outlined, first by describing the basic argument, as it is found in the expressivist literature (§1.2.1). Then, in §1.2.2, the problem will be examined in greater detail, by going back to its roots in a point made by Frege in his (1918-19). Finally, in §1.2.3, the implications of the Frege-Geach problem will be laid out, and the constraints and limitations that it places on expressivism will be made a bit more precise.

Before beginning, a few tedious but necessary remarks should be made about some of the basic terminology that will be used here. In the remainder of this paper, the predicates "right" and "wrong" will be called *moral predicates*, and well-formed sentences of English containing moral predicates will be called *moral sentences*. Non-moral declarative sentences will simply be called *descriptive* sentences, to reflect the intuitive idea that utterances of them express truth-apt propositions that describe a mind-independent physical world. For the purposes of simplicity, the class of all moral sentences will be called the *moral fragment* of natural language, and sentences belonging to this fragment will be said to be a part of moral discourse. On the other hand, ordinary descriptive discourse will be taken to be comprised of the class of descriptive sentences, and will be simply called the *descriptive fragment* of English. Much of this chapter will be concerned with making the distinction between moral and descriptive discourse more precise, and with laying down the key logical and semantic problems that arise from an expressivist interpretation of moral discourse.

1.1   Meta-ethical Expressivism

Consider the following moral sentence:

(1)   Lying is wrong

The meta-ethical expressivist takes a sincere utterance of (1) to express a non-cognitive attitude, bereft of truth conditions or propositional content. On this view, the semantic function of (1) is *not* to express a truth-apt proposition that describes or represents the world. Rather, its function is to express a desire-like attitude of *disapproval*: a feeling or moral sentiment, a preference or imperative, a higher-order attitude, or acceptance of a system of norms or a planning state.[1] In general, the meaning of (1) is explained in terms of its *use*, as a non-assertoric speech act, to *call* the act of lying wrong. While expressivists differ over the exact attitude that is expressed with this speech act, they uniformly agree that the semantic content of (1) differs markedly from that of the report

(2)   I disapprove of lying

for a sincere utterance of sentence (2) expresses a truth-apt proposition (relative to context), and provides a description of the mental state that the speaker is in.

These differences between (1) and (2) sharply demarcate the moral fragment of natural language from ordinary descriptive discourse, for any sentence containing a moral predicate is denied a descriptive, fact-stating, or representational role. These largely *negative* features form a cornerstone of the expressivist position, and can be captured more concisely in the following thesis:[2]

---

[1]See, for example, Ayer (1952), Hare (1969), Blackburn (1984), and Gibbard (1990, 2003).

[2]The term "nonfactualism" is due to Price (1994:132) and Boghossian (1990:159), and is similar to what Kit Fine calls "antifactualism" (Fine 2001:4). It is meant to be a general term that encapsulates noncognitivism, emotivism, projectivism, and quasi-realism.

Nonfactualism:

(i) Moral sentences are not truth-apt, they do not have truth-conditions, and they do not express propositions with truth-values.[3]

(ii) They do not state, describe or represent moral facts or states of affairs, nor do they aim to report on the world or refer to moral properties.[4]

(iii) They are not genuine assertions, and they do not express representational thoughts or beliefs.[5]

The semantic commitments captured by this thesis are driven largely by metaphysical considerations, by a reluctance to commit to any suspicious moral facts or properties. And it is this aversion to moral properties that results in expressivism being classified as either nonfactualist, irrealist, quasi-realist, or just generally anti-realist. Rather than doing an extensive taxonomic survey here, it may be more profitable to simply mark two general kinds of anti-realism that can be used to characterize expressivism:[6]

Anti-realism in ontology: there no specifically *moral* facts, objects or properties existing independently of the speaker.

Anti-realism in truth-value: moral sentences do not, and *cannot*, have objective truth-values, independent of the minds of speakers or the conventions of moral discourse.

The Nonfactualism thesis then makes expressivism both anti-realist in ontology and truth-value, for there are no moral properties for moral predicates to correspond to,

---

[3]Gibbard (1990:8), Ayer (1952:103, 107), Hale (1993:337, 340), Wedgwood (1997:73), Boghossian (1990:60), Stoljar (1993:81) and Sinnott-Armstrong (2000:678).

[4]Ayer (1952:89, 107), Boghossian (1990:160), Dreier (1999:563), and Gibbard (1990:8).

[5]Ayer (1952:103), Hale (1993:337), and Wright (1992:7)

[6]The notions of realism in ontology and truth-value are due to Shapiro (1993:456, 1997:37), who develops them for set theory. With a bit of tweaking, they apply just as well to moral discourse.

there are no facts to make moral sentences true or false, and there are no genuine moral assertions to be made.

But while expressivism denies that moral discourse has any substantial fact-stating or assertoric role, moral sentences at least *appear* to be able to do everything that ordinary descriptive sentences do, for they function, at the surface level of ordinary discourse, in much the same way that ordinary declarative sentences do. In practice, utterances of moral sentences are treated *as if* they are genuine assertions, *as if* they are truth-apt and express truth-conditions, and *as if* moral predicates correspond to properties. But the anti-realist and nonfactualist commitments of expressivism indicate that the syntactical and grammatical similarities between the moral and descriptive are deceptive, and that they obscure a fundamental semantic disconnect between the two domains. Crispin Wright (1992) puts it nicely, remarking that our ordinary realist tendencies are subject to a broadly Wittgensteinian diagnosis, in that we risk making an error;

> the error of being misled by merely superficial "grammatical" similarities between different regions of discourse, so that we are deceived into thinking that they are in the business of doing the same essential thing: stating facts, [or] expressing representational thoughts [...] What needs to be recognized is that a discourse may have all the overt trappings of genuine assertion, expression of belief and so on, yet in reality not be in the business of making statements, literally true or false, at all (Wright 1992:6).

So while moral and descriptive discourse appear identical at the surface level of ordinary discourse, below the surface the moral is severed from the descriptive by the Nonfactualism thesis. It clearly distinguishes the two domains by imposing negative constraints on moral sentences, by specifying what moral sentences *cannot* be committed to metaphysically, and what they *do not* do, semantically.

But there is a gap in the expressivists' theory, for denying that moral sentences have truth-conditions then requires that their meaning be explained non-standardly, without appealing to truth-conditions. So what is the primary semantic function of a moral sentence? According to expressivism, the function of an utterance of a sentence like "lying is wrong" is not to assert or describe, but rather to *express*:

Expression Thesis:

(i) The primary semantic function of an utterance of a moral sentence is to express a non-cognitive attitude: an emotion, desire, intention, feeling, or practical commitment.[7]

(ii) The meaning of a moral sentence is generally explained in terms of the *use* of the sentence, as a speech act, to express a non-cognitive attitude.[8]

The Expression Thesis is a largely *positive* position, in that it is designed to fill the semantic gaps that the Nonfactualism thesis leaves open. It provides an explanation of the meaning of a moral sentence *not* in terms of its truth-conditions, but in terms of a non-assertoric speech act of *calling* something right or wrong. And the primary function of this speech act is *not* to make an assertion that aims at truth, but rather to express a non-cognitive attitude of approval or disapproval that aims to motivate actions.

So what reasons are there for accepting this esoteric semantic theory? For one, it allows expressivism to avoid complicated reductivist or supervenience accounts of moral properties. And with no facts or properties playing a substantial role in the

---

[7]See Unwin (1999:337), Boghossian (1990:159), and Blackburn: "[they] express something more to do with attitudes, practices, emotions, feelings arising in contemplating some kinds of conduct, with goal seeking, with insistence upon normative constraints on conduct" (1988:504).

[8]The accounts in the literature differ slightly here, depending on the position being discussed. See, for example Sinnott-Armstrong (2000:679), and Unwin (2001:60).

theory, the expressivist can also skirt many of the metaphysical problems that plague realist programs in meta-ethics. In addition, the apparent action-guiding role of moral discourse is given an explanation *not* in terms of "queer" moral properties that have some spooky motivational force, but in terms of non-cognitive and desire-like attitudes. And finally, the distinction that expressivism draws between reporting and expressing makes it well-placed to solve problems with moral disagreement, thus giving it an advantage over subjectivist and indexical relativist theories.[9]

But while expressivism has certain theoretical advantages, the Nonfactualism and Expression theses engender a set of logical and semantic problems. Since the primary semantic function of moral sentences is to express a non-truth-apt attitude, it is unclear how they can be embedded into negated contexts, conditionals, and propositional attitudes, at least without having been issued something like a proper semantic import permit. And since moral sentences do not have truth-apt propositional contents, then absent any plausible semantic story, the classical accounts of validity, inconsistency, and logical consequence are bound to go by the board.

1.2   The Frege-Geach Embedding Problem

The severity of this situation is demonstrated by the Frege-Geach embedding problem, which is generally taken to be a decisive refutation of expressivism. The embedding problem finds its roots in a point that was made by Frege (1918-19) about the distinction between sense and force, and is based on a generalization of this point that Geach (1960, 1965) used to argue against a broad class of nondescriptivist theories.

---

[9]For more on these latter points, see especially Schroeder (2008:90-91).

In §1.2.1 I'll outline the general argument against expressivism, as it is recited in the expressivist literature. This I'll call the *moral modus ponens* argument. Then, in §1.2.2, I'll turn to Frege's *negation argument* in (Frege 1918-19), and show how this causes a problem for expressivism. Then I'll explain how Geach used Frege's result to establish a more general test to identify illicit forms of non-assertoric force – a test that is commonly called *Geach's test*. Finally, in §1.2.3, I'll ask if the embedding problem is the decisive refutation of expressivism that it is often claimed to be.

1.2.1 Moral Modus Ponens

Recall that expressivism attempts to explain the meaning of the moral sentence

(1) Lying is wrong

in terms of the non-assertoric speech act of calling lying wrong. So an utterance of (1) does not predicate a property of the action of lying, but rather functions to express an attitude of disapproval. This is largely unproblematic, and indeed it is theoretically advantageous, in the case of unembedded atomic moral sentences. But the grammatical structure of such sentences allows them to be freely embedded into more complex contexts: modal, epistemic, and propositional attitude contexts, negations, conjunctions, and conditionals like

(P2) If lying is wrong, then getting your brother to lie is wrong

And here expressivism encounters a problem, for the expressive force of "wrong" is suppressed when "lying is wrong" appears in the antecedent clause of the conditional (P2). There is no non-assertoric speech act of *calling* lying wrong when (1) appears in this embedded context, and so the predicate does *not* evince an attitude of disapproval. But since the meaning of an isolated occurrence of (1) is supposed to be given by the

very speech act that is suppressed in (P2), the meaning of "lying is wrong" *shifts* when it is embedded.

And this indicates that there's something that's not quite right about the expressivists' explanation of moral sentences. The full implications of this meaning shift are typically teased out by considering an argument of the following form:[10]

(F-G)   P1. Lying is wrong

           P2. If lying is wrong, then getting your brother to lie is wrong

∴ C. Getting your brother to lie is wrong

The argument is valid, bearing the surface form of *modus ponens*.[11] But the meaning of "lying is wrong" equivocates between (P1) and (P2), as the expressive force of "wrong" is canceled in the embedded context of the conditional. And since there is an equivocation of meaning, the argument is not valid.

The expressivist is placed in an awkward position; if **(F-G)** is valid, then the meaning of "lying is wrong" must remain invariant throughout the argument. But embedding "lying is wrong" into (P2) induces a meaning shift, so the meaning of "lying is wrong" does *not* remain invariant. So either moral sentences have a meaning distinct from expressive force, or **(F-G)** is not valid. If the first option holds, then

---

[10]This argument is standard fare in the expressivist literature, and is a variation of an argument given by Geach (1965:463). The notation **(F-G)** is meant, unsurprisingly, to stand for "Frege-Geach".

[11]It standard in the literature to make this assumption: it is "obviously" valid (Sinnott-Armstrong 2000:679, 680, 1993:300), (van Roojen 1996:313), (Stoljar 1993:91), "clearly" valid (Geach 1960:223), "intuitively" valid (Hale 1993:338), and simply "valid" (Blackburn 1984:190). The only dissenter to this view, to my knowledge, is Schueler (1988:496-7), who questions how the argument can be valid without the premises and conclusion being truth-apt.

the expressivists' explanation of the meaning of moral sentences is incomplete, and perhaps defective. If the second option holds, then the expressivist cannot explain even basic inferences containing moral sentences, and so has no account of the logical relations between moral sentences.

1.2.2 The Frege Point and Geach's Test

This is the standard *moral modus ponens* argument, as it is found in the contemporary expressivist literature. But it is only a gloss on a deeper problem that can be traced back to what is called the *Frege point*, which places restrictions on the placement of force-indicators in complex sentences:[12]

> **The Frege Point:** Force-indicators operate only on complete sentences, and never occur significantly within the scope of a logical or sentential connective.

The point is the linchpin of the *moral modus ponens* argument, and is based on a fundamental distinction that Frege draws between *propositions* and *judgments* — more specifically, between propositions and *assertoric* judgment.[13] A few words on this distinction might illustrate the significance of the Frege point.

For Frege, to assert a sentence A is to judge that the proposition expressed by A is true. This is represented formally by the assertion sign $\vdash$, which is composed of

---

[12]Geach (1965:449) calls his version of this the Frege point, but it also appears as *Frege's doctrine* in Michael Dummett: "a sign for force cannot occur within the scope of a sentential operator"(Dummett 1973:328, see also 316), and as the *Embedded Force Exclusion Principle* in Huw Price:"[f]orce modifiers cannot occur in embedded contexts"(Price 1994:137).

[13]Note that throughout this section I'll follow Sundholm (2002:571) and use the word 'proposition' in the place of Frege's word 'thought'. Furthermore, at times I'll conflate 'sense' and 'meaning', which is dangerous in other contexts (particularly when indexicals are involved), though I don't see that it will cause any intractable problems here.

a horizontal content stroke '–' and a vertical judgment stroke. Affixing the content stroke to A, giving '–A', is taken to mark the proposition expressed by A, or the propositional content of the sentence, an object which is apt for truth or falsity.[14] Completing the assertion sign by attaching the vertical judgment stroke to '–A', which gives '$\vdash A$', indicates that the proposition expressed by A is judged to be true.

While the propositional content of a sentence is apt for truth and falsity, a judgment is neither true nor false, and so it cannot fall under the scope of a truth-functional connective. This nicely separates propositional content from assertoric force, and truth-apt propositions from non-truth-apt judgments. And with this separation come the following three constraints, which underlie the Frege point:

c1 there can be no nested or iterated occurrences of $\vdash$.

c2 the truth-functional logical connectives operate only on propositions, and never on judgments.

c3 inferences operate only on judgments, and never on propositions.

One of the interesting features of these constraints is that they disallow any force-indicator, assertoric or otherwise, from appearing in an embedded context.

Call any context which does not allow a force-indicator to be embedded into it an *unasserted context*. Then, for example, to judge the propositions expressed by the individual sentences A and B to be true would be denoted by $\vdash A$ and $\vdash B$, respectively. Then the conditional $\vdash A \supset \vdash B$ is ill-formed, for the antecedent and

---

[14]This is not entirely accurate. As Dummett notes, (1973:314-316), Frege's interpretation of the content stroke differs between the *Begriffsschrift* and the *Grundgesetze*. In the former case, the stroke marks a proposition, and in the latter it marks a function from objects to truth-values. The distinction is unimportant here, particularly since the significance of the content stroke is questionable. Dummett even goes so far as to say that the stroke is "entirely superfluous."

consequent positions are unasserted contexts. But the judgment $\vdash (A \supset B)$, which marks that the proposition expressed by the conditional is true, is a well-formed judgment, and can enter into an inference. As another example, the inference

$$\frac{A \qquad A \supset B}{B} \, I_1$$

appears to be an instance of *modus ponens*, but strictly speaking it is ill-formed, as it does not operate on judgments, for the assertion signs have been suppressed.[15] This problem with inference I_1 can easily be remedied by simply affixing assertion signs to each complete sentence, as follows:

$$\frac{\vdash A \qquad \vdash A \supset B}{\vdash B} \, I_2$$

Then the inference I_2 is well-formed, relating a trio of judgments in the pattern of *modus ponens*.

Just as the assertion sign can be affixed to atomic sentences and conditionals, it can be affixed to negated sentences. According to Frege, a negated sentence like $\neg A$ expresses a negated proposition, and to assert $\neg A$ is to judge that this negated proposition is *true*. But since assertions judge propositions to be true, and propositions are paired off as being either true or false, it seems natural to also have a sign for denial, distinct from assertion, to mark the act of judging a proposition to be false. But Frege (1918-19) presents an argument, which I'll simply call the *negation argument*, against this. The remainder of this section details how this argument goes, and what the significance of it is. But before proceeding, I should remark that while it may

---

[15]It is common to suppress the assertion sign in inferences, though, as Martin-Löf remarks, "[c]onceptually, it would nevertheless be there" (1996:16). As an aside, Martin-Löf goes so far as to suggest that the assertion sign may be *identified* with a truth predicate.

appear that the current topic deviates significantly from the problem of embedding for meta-ethical expressivism, Frege's *negation argument* is in fact the primary source of the problem.

The *negation argument* goes like this. To show that there is no non-assertoric force of denial, Frege begins by supposing that, to the contrary, there *is* such a thing, and that it is represented by negation. That is, '*not*' is interpreted as an indicator of the non-assertoric act of denial, as a "polar opposite" (Frege 1918-19:383) to the assertion operator. Consider the following two negated sentences:

(3)   The accused was *not* in Rome at the time of the murder

(4)   The accused did *not* commit the murder

In both sentences negation takes wide scope, and marks the denial of a proposition. So, for example, with an utterance of (3) the negation indicates that the proposition expressed by "the accused was in Rome at the time of the murder" is judged to be *false*. The negation does not contribute to the *meaning* of "the accused was in Rome at the time of the murder," but rather marks a linguistic act, distinct from assertion, that conveys the force of denial. And this is designed to directly parallel the treatment of assertion. So far so good.

Now, it is an uncontroversial feature of negated sentences that they may be embedded in more complex contexts, as in the case of the conditional[16]

(5)   If the accused was not in Rome at the time of the murder,
      then he did not commit the murder

---

[16]Frege (1918-19:375) considers a slightly different sentence, of the form $\phi \supset \neg\psi$. I've changed it to one of the form $\neg\phi \supset \neg\psi$ in order to map on to the *moral modus ponens* argument.

Then, by the Frege point, neither the antecedent (3) nor the consequent (4) of the conditional are being denied, since with the act of asserting (5), the proposition expressed by the whole conditional is judged to be true, and neither the proposition expressed by (3) nor the proposition expressed by (4) are judged to be false. The negation 'not' does *not* function to mark the act of denial when the sentences (3) and (4) are embedded into the conditional, for the Frege point forbids force-indicators to fall under the scope of a logical or sentential connective. Instead, and most importantly, the negations contribute to the *meanings* of the antecedent and the consequent of (5).[17] That is, the meanings of both (3) and (4) are modified, or *shift*, when they are embedded into the antecedent and consequent (respectively) of a conditional, since in those contexts the negation functions as a sentential (or propositional) operator, and not as a force-indicator.

So just as the meaning of the moral sentence "lying is wrong" shifted when it was embedded into the conditional (P2), the meaning of the negated sentence (3) shifts when it is embedded into (5). And in much the same way that the meaning shift of the moral sentence (1) caused the inference **(F-G)** to go awry, so will any (apparent) instance of *modus ponens* containing a negated sentence. Frege shows this by comparing the two inferences (1918-19:379-380):[18]

$$\frac{\neg\phi \qquad \neg\phi \supset \neg\psi}{\neg\psi}\,\text{Neg} \qquad\qquad \frac{\phi \qquad \phi \supset \psi}{\psi}\,\text{MP}$$

A quick comparison of these inferences shows that they both appear to be of the same logical form, or as Frege says, they "proceed in the same form"(1918-19:380),

---

[17]See Dummett (1973 :316-317) or Price (1994:136) for a similar analysis.

[18]I've modified the second, **(MP)**, so as to line up a bit with the *moral modus ponens* argument. Frege's original is of the form $\phi \supset \neg\psi$; ϕ; therefore $\neg\psi$ (1918-19:380). To line up with the conditional (5), read ϕ as "the accused was in Rome at the time of the murder," and read ψ as "the accused committed the murder."

namely *modus ponens*. That much is clear. However, there is a sharp and crucial difference between these two inferences, which is not readily apparent given the current formalizations, but which becomes evident when the relevant force-indicators have been formally represented.

Up to this point, I've followed Frege and marked the act of judging false with the negation operator '$\neg$'. But given that the force of denial is canceled in embedded contexts, the distinction between content and force can be made more explicit by following a suggestion of Geach's (1965:455), and using Łucasiewicz' force-indicator $\dashv$ to denote the act of denial. So $\dashv A$ marks the act of judging the proposition expressed by A to be false, while the assertion $\vdash \neg A$ indicates that the proposition expressed by $\neg A$ is true. Then the sentence (3) would be regimented as $\dashv \phi$, and (4) would just be $\dashv \psi$. And by taking the Frege point into account, the conditional (5) would be formalized as $\vdash (\neg\phi \supset \neg\psi)$. Then **(Neg)** and **(MP)** can be written as:

$$\frac{\dashv \phi \qquad \vdash \neg\phi \supset \neg\psi}{\dashv \psi} \; (\text{Neg}^\vdash) \qquad\qquad \frac{\vdash \phi \qquad \vdash \phi \supset \psi}{\vdash \psi} \; (\text{MP}^\vdash)$$

The additional detail here indicates what exactly is being asserted and denied, and what the scope of each force-indicator and sentential operator is.

So what's the problem? Well, in **(MP$^\vdash$)**, the meaning of ϕ is the same for each occurrence of it in the inference. And it is this invariance in the meaning of the sentence that, in part, explains the validity of the inference. But a quick look at **(Neg$^\vdash$)** shows that there is no similar invariance in the meaning. The meaning of the unembedded ϕ differs from the meaning of the $\neg\phi$ embedded in the antecedent of the conditional, for the latter expresses a negated proposition. And as a result, the inference **(Neg$^\vdash$)**, to use Frege's words, "cannot be performed in the same way as before" (1918-19:384). That is, by visually comparing the two inferences, it becomes

clear that they do not proceed according to the same logical principles. The inference $(\mathbf{Neg}^\vdash)$ is not an instance of *modus ponens*, despite initial appearances to the contrary.

Here's what's going on. If there were a new non-assertoric force for denial, then the inference $(\mathbf{Neg}^\vdash)$ would *not* proceed in "the same way" as *modus ponens*, despite the fact that it seems to, at least based on its surface form. And the only way to make $(\mathbf{Neg}^\vdash)$ follow the same logical principles as *modus ponens* would be, according to Frege, to have the unembedded premise $\dashv \phi$ express the same proposition as $\neg\phi$. But this would involve "separating negation from the act of judging, [...] and uniting negation with the [proposition]" (Frege 1918-19:384). In other words, if we want $(\mathbf{Neg}^\vdash)$ to be an instance of *modus ponens*, then negation *can't* mark an act of denial and it *must* be a sentential operator.

It is important to note that Frege does not explicitly deny that $(\mathbf{Neg}^\vdash)$ is valid, only that when it is held up against $(\mathbf{MP}^\vdash)$, it becomes clear that the two patterns don't line up. So it seems, at first glance, that there is an easy way out here; simply postulate a *new* inference pattern to accommodate $(\mathbf{Neg}^\vdash)$, and specify some *key semantic concept* that is preserved in the inference.[19] However, Frege disallows this move, on the grounds that it would require more resources to explain the apparent validity of $(\mathbf{Neg}^\vdash)$, beyond those already available to explain ordinary *modus ponens*. Similarly, postulating a new non-assertoric force for denial would require three distinct devices, (i) assertion, (ii) negation, and (iii) a sign for denial, whereas Frege's preferred method would involve only two: (i) assertoric force and (ii) negation. Here's what Frege says, referring to the "economy" gained by postulating fewer logical symbols

---

[19]The term 'key semantic concept' is due to Sundholm (1986:473), who describes it as the key feature of a theory of sense that both gives content to every sentence, and that explains compositionality ("iterated unboundedness"). Typically, *truth* is taken to be the key semantic concept in a theory of meaning, but it can also be given by e.g.; *proof.*

and forms of inference:

> Such economy always shows that analysis has been pushed further, which
> leads to a clearer insight [...] If we *can* make do with one way of judging,
> then we *must* (1918-19:385).

And it is this preference for economy that is supposed to disallow any additional
inference patterns or force-indicators to explain $(\mathbf{Neg}^{\vdash})$ and denial.[20]

At first glance, Frege's negation argument might appear to be relatively harm-
less, and tangential to the expressivist project in meta-ethics. However, just as it is
shown that negation cannot mark the act of denial, so too can any other alleged force
indicator be tested to see if it is legitimate. As Dummett says, the Frege point, if
it is correct, "appears to constitute a powerful method for detecting spurious claims
to have identified a new kind of force, or form of linguistic act" (Dummett 1973:327).
And this is something that the expressivist should be concerned about, since the
meaning of a moral sentence like "lying is wrong" is typically explained by its use as
a non-assertoric speech act to *call* lying wrong. That is, "wrong" is a force indicator
in disguise as a predicate.

Let's see how this goes by using the template of the *negation argument* to argue
against the treatment of "wrong" as a new kind of non-assertoric force. First, suppose
that the expressivist is correct, and that the predicate "wrong" marks a linguistic act
distinct from assertion that conveys an expressive force of disapproval. Then "wrong"
does not contribute to the propositional content of the sentence "lying is wrong", but
instead serves to indicate a new kind of non-assertoric and *expressive* force.

---

[20]As we'll see in Chapter 2, the sophisticated expressivist rejects this claim, and is happy to
postulate a special moral conditional that subsumes the material conditional, and a more general
inference pattern that subsumes *both* the moral and descriptive versions of modus ponens.

But then we're confronted with the familiar problems that stem from the Frege point. Since the expressive force of "lying is wrong" is canceled in unasserted contexts, lying is not called wrong when (1) appears in the antecedent of the conditional (P2). Again, there is a meaning shift induced by the embedding.[21] And again, Frege's comparison method will reveal that the inference **(F-G)** does not proceed in the same way as *modus ponens*. To make this more explicit, let's look at the forms **(F-G)** and *modus ponens* in more detail:[22]

$$\frac{W(l) \qquad W(l) \supset W(b)}{W(b)} \text{ F-G} \qquad\qquad \frac{\phi \qquad \phi \supset \psi}{\psi} \text{ MP}$$

Again, appearances are deceptive, for the basic form obscures the implicit force-indicators. So let '$!_w$' be a sign for expressive force that marks the act of calling something wrong. Then, since the expressive force of "wrong" is canceled in embedded contexts, the conditional (P2) is asserted overall, and is properly written as '$\vdash W(l) \supset W(b)$'. Then the inferences look like

$$\frac{!_w\, l \qquad \vdash W(l) \supset W(b)}{!_w\, b} \text{ F-G}^\vdash \qquad\qquad \frac{\vdash \phi \qquad \vdash \phi \supset \psi}{\vdash \psi} \text{ MP}^\vdash$$

A quick comparison of these forms shows that **(F-G$^\vdash$)** deviates from *modus ponens*, since the meaning of the first premise in **(F-G$^\vdash$)** is different than that of the antecedent $W(l)$.[23] Then, following Frege's comparison method, this means that the two inferences do not proceed by the same logical principles, and so "wrong" must be

---

[21]Note that this doesn't exactly mirror the *negation argument*, since there the embedded negations were taken to contribute to the propositional content of the antecedent and consequents. Here, however, it is not clear if the embedded occurrences of the predicate "wrong" operate in the same way, since they are not supposed to be predicates with extension.

[22]Here W = "wrong", l = "lying" and b = "getting your brother to lie".

[23]Again, it is left open precisely *how* the meanings of $!_w l$ and $W(l)$ differ, since the latter is unanalyzable if the expressivist does not explicitly commit to treating embedded occurrences of the predicate 'W' as truth-functional and extensional.

an illicit force indicator. The meaning of "lying is wrong" cannot be given by the act of calling lying wrong, by expressing an attitude of disapproval. Expressivism may be in trouble.

Again, note that at no point was it explicitly claimed that the inference **(F-G)** is not valid. Rather, all that was shown is that a comparison of **(F-G)** and *modus ponens* shows that the former cannot be an instance of the latter, and that the only way to fix this discrepancy would result in an 'uneconomical' theory. This differs from the standard *moral modus ponens* argument against expressivism, both in strategy and in conclusion. The standard argument against expressivism does not use Frege's comparison method to show that **(F-G)** is not a case of *modus ponens*, but instead makes the stronger claim that an equivocation in the meaning of (1) is enough to show that **(F-G)** is *not valid*.[24]

This latter approach, which has become standard in the expressivist literature, follows a template devised by Geach (1960, 1965), who leverages Frege's *negation argument* to devise a more general test that is designed to identify and eliminate any illicit non-assertoric force-indicators. This test, which is commonly called *Geach's test*, preys on what Geach calls a "radical flaw" (1960:222) in the non-descriptive and expressive interpretation of a predicate; the flaw of conflating the act of *calling* a thing "wrong" with *predicating* "wrong" of it. Speaking of the predicate "bad", in place of "wrong", Geach refers obliquely to the Frege point by saying

---

[24]Note, however, that not all standard expressivist accounts of the embedding problem eschew the comparison method. Blackburn (1988) refers to the comparison method (without calling it that), saying of **(F-G)** that "[i]t is modus ponens, for all that. Or if we say that it is not, then we have no effective procedure for telling when anything is" (1988:502). Note how this differs from his earlier account of the inference in his (1984), where he simply says that **(F-G)** is "a valid argument" (1984:190), and makes no reference to Frege's comparison method, instead relying on an equivocation in the meaning of "wrong" to make the problem go through.

> [a] term "P" may be predicated of a thing in an *if* or *then* clause, or in
> a clause of a disjunctive proposition, without the thing's being thereby
> called "P." [...] [T]o say, "If gambling is bad, inviting people to gamble
> is bad" is not to *call* either gambling or invitations to gamble "bad."
> [...] [P]redications of [...] "bad" in *if* or *then* clauses, or in clauses of a
> disjunction, are just ignored (Geach 1960:223).

This part of the story is familiar. But Geach departs from Frege by dispensing with the
comparison method, and going straight to a claim about the validity of an argument
of the form "if gambling is bad, inviting people to gamble is bad; gambling is bad;
therefore inviting people to gamble is bad". Speaking of the embedded occurrences
of "bad" in the *if* and *then* clauses, Geach says that

> [o]ne could not write off such uses of the terms, as calling for a differ-
> ent explanation from their use to *call* things true or bad; for that would
> mean that the argument ["if gambling is bad, inviting people to gamble
> is bad; gambling is bad; therefore inviting people to gamble is bad"] con-
> tained a fallacy of equivocation, whereas [it is] in fact clearly valid (Geach
> 1960:223).

The requirement that valid arguments not be subject to equivocation is not explicitly
stated by Geach here, though he remedies this somewhat in his (1965) by linking the
validity of *modus ponens* with equivocation:[25]

> [...] the two occurrences of ["gambling is bad"], by itself and in the "if"
> clause, must have the same sense [or meaning] if the *modus ponens* is not
> to be vitiated by equivocation [...] (Geach 1965:463).

---

[25] Still, however, the precise connection between equivocation and validity is left unexplained by
Geach. Price takes Geach to be making an implicit appeal to what he calls *sense identity*, which
says that *modus ponens* is "valid only if the [first] premise has the same sense (or expresses the same
thought) as the antecedent of the conditional premise" (Price 1994:134).

So Geach's generalization of Frege's *negation argument* makes no explicit appeal to a comparison between problematic inferences and genuine cases of *modus ponens*, but instead relies on the Frege point to show that, assuming a non-descriptive reading of a predicate, embedding that predicate into an unasserted context engenders an equivocation in the explanation of the *use* or *sense* of the predicate.[26] This, according to Geach, is enough to invalidate the argument and discredit nondescriptivist interpretations of "bad" or "wrong". And, crucially, this result is generalizable to argue against a non-descriptive and expressive reading of *any* predicate whatsoever.

1.2.3 Some Remarks on the Scope of the Problem

The effectiveness of both Geach's test and the *negation argument* hinge on the Frege point. So the expressivist might be inclined to question the legitimacy of the point, perhaps in an effort to deflect the embedding problem by finding an expressivist-friendly force-indicator that could fall under the scope of the logical connectives.

Michael Dummett (1973:327-48) does something along these lines, looking for various kinds of force that can appear embedded in complex sentences. He finds some exceptions to the Frege point, but he is unable to give a systematic and general schema that provides anything more than an *ad hoc* and case-by-case search for these outlying exceptions. As a result, he is led to conclude that all exceptions to the Frege point are merely "superfluous" (1973:340), and do not undermine it or "effect its essential soundness"(1973:348). So much for that strategy.

---

[26]Clearly 'use' and 'sense' are not interchangeable. In his (1960), Geach makes no explicit mention of 'meaning' or 'sense', instead cashing the problem out in terms of the way that a predicate is *used*: in unembedded contexts, "gambling is bad" is *used* to condemn gambling, but when embedded in a conditional no such use is made. In his (1965), however, he reverts to talk of 'sense' in the place of 'use', which leads more naturally into the old *moral modus ponens* argument.

So if the embedding problem can't be solved by finding appropriate exceptions to the Frege point, then should it be taken to be a decisive refutation of expressivism? Both Price (1994:133) and Hale (1993:338) argue that the force of the Frege-Geach argument may be overstated, and that at worst it functions to place certain minimal, though still serious, constraints on the complexity of moral discourse. Hale in particular takes the embedding problem to show that a satisfactory speech-act analysis of the expression of a non-cognitive attitude *must* have have a degree of complexity "paralleling" that of assertion (1993:338). This suggests that moral judgment likely requires theoretical devices to play the role of *truth* and *proposition* in the semantic theory, so that instances of moral *modus ponens* do not succumb to an equivocation of meaning.

Hale (1993:338-339) refines this idea by saying that while a sincere assertion of a sentence *expresses* a proposition and *registers a commitment* to its truth, a viable expressivist analogue would take a sincere affirmation of "lying is wrong" to *present* an attitude of disapproval, and to *register a commitment* to the attitude. Modulo differences between *expressing* a proposition and *presenting* an attitude, a distinction that Hale leaves unexplained, the expressivist version of assertion would then appear to replace truth-apt propositions with non-cognitive attitudes, and judgments with the more general act of registering a commitment to an object. Then, in the case of descriptive sentences, a commitment would be registered to the truth of a proposition, while in the moral case a commitment would be registered to an attitude. So in affirming the conditional (P2), a speaker would *present* an attitude of disapproval towards lying, but she wouldn't *register* commitment to it. And this is designed to mirror the standard case where asserting a descriptive conditional involves *expressing*

the proposition associated with the antecedent, but it does not involve the act of *judging* that proposition to be true.

As I've read Hale, it appears that the point of this is to establish a degree of complexity in the speech act of expressing a non-cognitive attitude that allows a certain semantic connection to be made between asserted and unasserted occurrences of moral sentences. That is, solving the embedding problem would require, at least, that there be something essential in the meaning of (1) that tags along with the sentence into unasserted contexts, unlike the force-indicators that are prevented, by the Frege point, from making a similar trip. As Hale nicely puts it, at a minimum the Frege-Geach problem "debars [the expressivist] from offering wholly disconnected accounts of the semantic roles of asserted and unasserted components" (1993:338).

This idea is similar in theme to one discussed by Price (1994:135-136) and Hare (1970:9), which is based on an apparent loophole in the Frege-Geach argument that was discovered by Searle (1969). The loophole consists in observing that the meaning of a moral sentence like "lying is wrong," when it appears in an unasserted context, need not be explained entirely by the actual performance of the speech act of calling lying wrong. Though Searle is concerned with general speech act analyses, he in effect thinks that expressivists

> do not need to show that every utterance of [(1)] is a performance of [the act of calling lying wrong], but rather they need only to show that literal utterances which are not performances of the act [of calling lying wrong] stand in a relation to performances of [the act of calling lying wrong] in a way which is purely a function of the way the sentences uttered stand in relation to the standard indicative sentences, in the utterance of which the act is performed (Searle 1969:138).

That is, there is no strict requirement that the meaning of embedded and unembedded occurrences of (1) be exactly identical, but rather what is required is that a more nuanced semantic connection between embedded and unembedded occurrences be established.[27] Hare explains this in terms of a "transformation, whose form we understand" (1970:14) between an unembedded occurrence of a sentence, where the original speech act is performed, and an embedded instance of the sentence. And as Price says, speaking of negation in an unasserted context, "[i]t is enough that its contribution to the meaning of the containing context should depend on the fact that it does signal a denial, when used [unembedded]" (Price 1994:138). And this applies just as well to moral sentences, suggesting that a refinement in the account of expressive meaning might, along the lines suggested by Hale, Price and Searle, offer the most promising solution to the embedding problem.

So it seems that, at a minimum, the meaning of (1) has to be explained in such a way that embedding it in the conditional (P2) does not violate the Frege point. But this is only the first step in a larger project of giving a compositional semantics and an account of logical consequence for the moral fragment of natural language. The problem is that the expressivist wants to accommodate the apparent factualist surface phenomena of moral discourse, and explain how moral statements *appear* to function just like descriptive statements, and how they *appear* to be truth-apt and representational. The rub is that this explanation cannot involve any robust notions of *fact*, *property*, *truth* or *truth-condition*, and can make no direct appeal to ordinary truth-functional logic. What the expressivist demands, then, is a semantic theory that appears, *on the surface*, to integrate seamlessly into a denotational truth-conditional

---

[27]Note that this may end up conflicting with the *sense identity* condition (see ft.25) that Geach requires for his version of the embedding of the argument to go through.

semantics for ordinary descriptive language. But, *below the surface*, so to speak, the semantic machinery must operate without availing itself of any overtly realist concepts.

The next chapter explores the possibility of such a program by examining two prominent expressivist theories, Simon Blackburn's quasi-realism (§2.1) and Allan Gibbard's norm-expressivism (§2.2). My main concern the next chapter is rather narrow, in that I am primarily concerned with how these theories address two key problems: the first problem is how to give an explanation of the meaning of the sentence (1) in such a way that it can be embedded into the conditional (P2) without running afoul of the Frege point. I'll call this the problem of *embedding*. The second problem is how to explain the apparent validity of **(F-G)** without appealing to a notion of truth that violates the expressivists' commitments to anti-realism in ontology and truth-value. I'll call this the problem of *logic*.

Chapter 2

Sophisticated Expressivism

Confronted with the problems raised by Geach and Frege, contemporary expressivist theories have attempted to deflect the embedding problem by providing a more complex account of the meaning of moral sentences. This is generally done by explaining the meaning of a moral sentence in terms of the mental state that it is conventionally used to express. Gibbard (2003) describes this strategy as *changing the question*:

> Don't ask directly how to define 'good' [...] Ask what *states of mind* ethical statements express [...] to explain the meaning of a term, explain what states of mind the term is used to express (Gibbard 2003:6-7).

This general idea is familiar from the Expression Thesis, but here the expressivist goes one step further and extends it to apply to *every* declarative sentence in natural language, resulting in what Wedgwood (2007:35) calls a *psychologistic semantics*. On this view, the primary semantic function of *every* declarative sentence, whether it be descriptive or moral, is to express a mental state of some sort. And it is this mental state that is said to confer upon the corresponding sentence its meaning.

The crucial distinction between descriptive and moral discourse is then cashed out in terms of the *kind* of attitude that the sentences from each domain are conventionally used to express: genuine assertions of descriptive sentences are taken to express truth-apt and representational beliefs, while moral judgments are taken to express some other sort of non-cognitive and desire-like attitude. And it is the distinction between representational beliefs and motivating desires that draws the line between descriptive and moral discourse.

Then the next step in the expressivists' strategy is to construct a compositional semantics for attitudes that functions in the same way as an ordinary truth-conditional semantics. The aim is to provide a semantic theory where the attitude expressed by a complex sentence is determined recursively in terms of the attitudes expressed by its parts. Here's Gibbard on the project:

> For sentences that express truths or falsehoods, we have a rich legacy of accounts that tell us how this happens: Tarski and his successors tell us how the truth conditions of a sentence depend on the truth conditions of its components. Can we do anything of the sort with an expressivistic theory? A normative sentence, the expressivist says, expresses a state of mind; its meaning is explained not by giving its truth conditions but by telling what state of mind it expresses. When a normative term appears in a complex context, can we still say what state of mind is being expressed? Can we give a systematic account of how the state of mind a complex normative sentence expresses depends on the states of mind that would be expressed by its components alone? (Gibbard 1990:92)

It is the burden of expressivism to answer these questions in the affirmative, to provide a compositional semantics for *attitudes* that mirrors standard truth-conditional semantics. So expressivism has the daunting task of explaining how the mental state

expressed by a compond sentence is a function of the mental states expressed by its subsentences. The expressivists' idea is to replace *propositions* with *attitudes*, and then give a logic and semantics for these objects. A chart might make this a bit more explicit:

	(Old) Fregean picture	(New) Expressivist picture
logical constants operate on	propositions	attitudes
inferences operate on	judgments	???

For lack of a precise definition, the expressivist analogue of judgment is left open here, but by the end of this chapter some candidates will be available. The first row, however, shows how this new approach might be able to deflect the Frege point. To see how this works, let's try to embed (1) into the following conditional, a *mixed* moral and descriptive sentence[1]

   (6)   If lying is wrong, then my mom will get angry at me

Just as above, an utterance of (1) is taken to express an attitude of disapproval, a non-cognitive and motivating desire-like attitude. But here the descriptive consequent does *not* express a proposition, but an attitude, a representational *belief* (relative to context). And the conditional (6) is also taken to express an attitude overall, which is determined recursively in terms of the attitude of disapproval that (1) is conventionally used to express, and the *belief* that the speaker's mother will be mad at them. The expressivist's goal is then to give a general logic and semantics for *attitudes*, one that subsumes both descriptive and moral discourse.

---

[1]This works equally as well for the familiar case of embedding (1) into (P2), but the current mixed sentence is more difficult (and more interesting) to explain because it mixes sentences from the moral and descriptive domains.

Of course, all of this depends on giving a proper logic and semantics for attitudes. And that is the subject of this chapter, which looks at two attempts to do just this. The structure of the chapter is as follows: section §2.1 details two separate attempts made by Simon Blackburn to give a logic and semantics for his quasi-realist brand of expressivism. In §2.1.1, Blackburn's (1984) *slow track* attempt to provide an expressive interpretation of the conditional and *modus ponens* will be outlined. Then, in §2.2.2, a more sophisticated proposal, Blackburn's (1988) *logic of attitudes*, will be explained. Here, the basic non-cognitive attitudes of approval and toleration are formalized by the intensional operators $H!$ and $T!$, which are expressive variants of the ordinary deontic operators of obligation and permission. Blackburn's proposed solution to the Frege-Geach embedding problem involves an expressive interpretation of deontic logic, and while this seems initially promising, in §2.1.3 it will be shown that the logic of attitudes faces two severe problems that threaten to undermine it.

The topic of §2.2 is Allan Gibbard's norm-expressivism, which provides an even richer and more complex account of the attitudes of approval and disapproval. By interpreting norm-expressivism in a modified possible worlds framework, a *normative logic*, Gibbard is able to present convincing and elegant solutions to the problems of embedding and logic. And at first glance, it appears as though he has solved the Frege-Geach embedding problem. However, in §2.2.1, it will be argued that Gibbard's formalism is only masquerading as an expressivist theory, and is actually an indexical theory in disguise. And, as a result, normative logic collapses into a cognitivist or realist meta-ethical position.

2.1 Blackburn's Quasi-Realism

In his *Spreading the Word* (1984) and "Attitudes and Contents" (1988), Simon Blackburn develops a logic of attitudes for his *quasi-realist* program in meta-ethical expressivism. Quasi-realism aims to explain how an expressivist can speak *as if* they were a realist about moral discourse, and *as if* moral sentences are truth-apt and descriptive, even though their primary function is expressive, and not descriptive. This involves what Blackburn calls *projectivism*, where

> [...] we *project* an attitude or habit or other commitment which is not descriptive onto the world, [where] we speak and think as though there were a property of things which our sayings describe which we can reason about, know about, be wrong about, and so on (Blackburn 1984:170-1).

Projectivism is designed sidestep the problems associated with committing to moral facts or truth-apt descriptions of moral states of affairs. Moral, or *evaluative*, sentences like (1) are taken to involve an *evaluative commitment* to a non-cognitive attitude, to evincing an attitude of disapproval towards an act. *Descriptive commitments*, on the other hand, are simply said to be beliefs or assertive judgments that a proposition is true.[2] This distinction then makes Blackburn's expressivism susceptible to the Frege-Geach embedding problem, a proper solution to which requires, according to Blackburn, an answer to the general question of "what we're up to" when we're using logical connectives to pair descriptive sentences with moral sentences. And this

---

[2]When contrasting evaluative with descriptive commitments, Blackburn variously associates descriptive commitments with beliefs, assertions, propositions, judgments with truth-conditions, and "asserting that some state of affairs obtains" (1984:167). And he says that commitments, whether they be descriptive or evaluative, are "accepted" (1984:192) by speakers, and that they "express" a speakers' attitude (1984:188,190).

amounts to explaining how moral sentences can fall under the scope of the logical connectives but still retain an expressive interpretation.

Blackburn's proposal is to "expand the way we think of [the connectives]" (1984:190), to provide a more general and non-truth-functional interpretation of the logical connectives in terms of *commitments*. So instead of operating on truth-apt propositions, Blackburn takes the connectives to operate on commitments, on non-cognitive attitudes and beliefs. Referring to the specific case of conjunction, Blackburn says that

> [we] might explain the semantic function of 'and' like this: it stands between two sentences to make one large sentence out of them; the large sentence is true if and only if each smaller one is true. Otherwise it is false. Now this little semantic theory fits badly, initially at any rate, with the occurrence of the evaluation as a conjunct [...] We could instead say something like this: 'and' links commitments to give an overall commitment which is accepted only if each component is accepted (1984:191-192).

So, according to Blackburn, what we're up to when we're using binary connectives is we're linking commitments, whether they be beliefs or non-cognitive attitudes, to recursively generate more complex commitments. And this is the starting point for a non-truth-conditional commitment semantics that captures how moral sentences can be used *as if* they were descriptive, *as if* they entered into logical relations and described moral facts in the world, and *as if* they expressed truth-apt propositions.

2.1.1   The Slow Track: Honest Toil Over Theft

Blackburn's (1984) proposed solution to the Frege-Geach problem focuses first on providing an account of conditional embedding of commitments, and then aims to address each of the remaining problematic embedded contexts in turn. This is a *slow*

track solution that begins with the conditional, and that attempts to explain the apparent validity of the moral *modus ponens* argument form **(F-G)** in terms of a general logic and semantics of commitments. Blackburn (1994:193-4) gives a rough outline of this by defining a first-order language E_{ex} with no evaluative predicates, and with the following devices:

d1 Operators $H!$ and $B!$ that operate on descriptions of actions. The function of these devices is to express a particular attitude, $H!$ expressing an attitude of approval, and $B!$ expressing an attitude of disapproval. So $B!(\text{lying})$ expresses an attitude of disapproval towards the act of lying.

d2 A device '$|...|$' for referring to commitments. So $|B!(\text{lying})|$ refers to a disapproval of lying, and $H!|B!(\text{lying})|$ expresses a higher-order attitude of approval of disapproval of lying.

d3 A symbol '$\Rightarrow$' that denotes an *involvement* relation, a binary relation on descriptions of attitudes.[3] So the involvement of the attitude of disapproval expressed by $B!(\text{lying})$ in the attitude expressed by $B!(\text{getting your brother to lie})$ is represented by $|B!(\text{lying})| \Rightarrow |B!(\text{getting your brother to lie})|$. This is designed to represent the idea that the attitude referred to by $|B!(\text{getting your brother to lie})|$ *follows upon* the attitude referred to by $|B!(\text{lying})|$.[4]

This gives only a basic sketch of the language E_{ex}, but still there is enough here to demonstrate how Blackburn addresses the problems raised by the Frege-Geach argument **(F-G)**. Consider the following sequence of sentences, which are the E_{ex} analogs of sentences (P1) and (P2):

(P1ʹ) $B!(\text{lying})$

(P2ʹ) $H!(|B!(\text{lying})| \Rightarrow |B!(\text{getting your brother to lie})|)$

---

[3]Note that I am importing notation from Blackburn's (1988:507) here. In his (1984), the symbol used is ';', and is designed to also cover a 'coupling' relation that is similar to conjunction.

[4]See Blackburn (1984:195) and Hale (1986:73, 1993:343) for more on the '*follows upon*' relation.

Let's look at what's going on here. The sentence (P1′) expresses a disapproval of lying, a (projective) relation between an attitude and an act. The second sentence (P2′) is operated on by the $H!$ operator, and so expresses an approval of an involvement of an attitude in an attitude. In particular, (P2′) expresses approval of the involvement of the disapproval of lying in the disapproval of getting your brother to lie. So the conditional expresses a commitment overall, a second-order attitude of approval that is determined by referring to the attitudes that *would* be expressed by its components were they in unasserted positions. And with this, the Frege point is deftly sidestepped, for even though the attitude of disapproval expressed by $B!$(lying) is suppressed when it is embedded, that attitude is still referred to when $B!$(lying) is operated on by the device '$|...|$,' and when $|B!$(lying)$|$ is embedded into the antecedent position of (P2′). This is, in effect, an exploitation of Searle's (1969:138) loophole to explain the embedding problem for conditionals.

Now let's see how Blackburn's proposal fares with the problem of logic. So return to the original argument **(F-G)**, and consider the conclusion (C), that getting your brother to lie is wrong. On Blackburn's proposal, if one held (P1′) and (P2′), then, the story goes, they *must* also disapprove of (C). For if they did not disapprove, they would simultaneously disapprove of getting their brother to lie while *not* disapproving it. And this would amount to what Blackburn calls a 'clash of attitudes,' (1984:195), an expressivist analogue of the logical notion of inconsistency. It is designed to indicate that a disapproval of (C) *should not* follow from (P1′) and (P2′), and that to avoid a clash of attitudes, anyone committed to (P1′) and (P2′) *must* also be committed to

(C′) $B!$(getting your brother to lie)

And this is enough, for Blackburn, to declare that the inference from (P1′) and (P2′) to (C′) is justified. With a judicious application of the sentential operators $H!$ and $B!$, the apparent validity of **(F-G)** is given an explanation.

The virtue of this approach lies in its generality. By abstracting to the level of commitments, whether they be beliefs or attitudes, Blackburn is able to provide a surrogate for implication that licenses the use of evaluative statements *as if* they expressed a propositional content and were assigned truth-conditions. And at this level of abstraction, avoidance of a clash of attitudes is, according to Blackburn, sufficient to license the inference **(F-G)**.

However, with this generality comes a lack of precision, for the sharp truth-functional explanations of the conditional and logical inconsistency are supplanted by the non-logical involvement relation '$\Rightarrow$' and the more general notion of a 'clash of attitudes'. Hale (1986:73) in particular questions whether these new non-logical concepts are enough to properly explain implication and inconsistency. For in foregoing any "descriptive" or truth-functional explanation, both the *follows upon* relation that characterizes '$\Rightarrow$', and the conflicting commitments that result in a clash of attitudes, fall short of their logical counterparts.

Schueler (1988) argues similarly, charging that Blackburn's 'clash of attitudes' conflates two distinct notions, *logical* inconsistency and *pragmatic incoherence*. And once these two notions are untangled, it becomes clear that a clash of attitudes is a *weaker* notion than logical inconsistency. So while abstracting to the level of attitudes allows Blackburn to adhere to the main tenets of the expressivist and quasi-realist positions, doing so might make the resulting theory lose much of its logical bite.

2.1.2 The Fast Track: A Logic of Attitudes

Blackburn's 'slow track' attempted to give a semantics for the conditional in terms of the non-logical involvement relation '⇒', and exploited the fact that at a certain level of abstraction, at the level of *commitments*, the logical behavior of moral and descriptive sentences seems to coincide. The intent of this was to render a plausible account of valid inference (in the form of *modus ponens*) for moral statements. And all of this culminated in something resembling a proof of the validity of **(F-G)** by appealing to the notion of a clash of attitudes.

In "Attitudes and Contents" (1988), Blackburn shirks several aspects of the slow track approach, and instead follows a more comprehensive 'fast track' that attempts to account for most, if not all, of the problematic embedded contexts simultaneously, thus arriving at a solution "in better style"(and who would wish to deny him that?). Fast track quasi-realism is designed to give a truth-surrogate, a (nonrepresentational) key semantic concept that functions to "regulate evaluative discourse"(1988:505). Says Blackburn,

> the point is to *earn* our right to propositional forms - including the use of a truth predicate. If this is done, any conventional concept of validity tags along - there is a level of analysis at which modus ponens and the rest are no different when their components are evaluative and when they are not (Blackburn 1988:505).

While Frege's parsimony disallowed a form of inference distinct from *modus ponens* to explain the inference **(F-G)**, on the grounds that this would result in an "uneconomical" theory, Blackburn attempts to find a more general inference pattern that subsumes both ordinary *modus ponens* and the problematic inference **(F-G)**.

Blackburn's strategy is to give additional logical structure to the basic attitudes of approval and disapproval by framing his brand of expressivism in Hintikka's (1969) deontic logic. This is designed to give expressive content enough structure to dissociate it from non-assertoric force, making the resulting theory well-placed to deflect the embedding problem. The challenge is to develop this into a general *logic of attitudes*, and all without deploying any realist or representational machinery.

In Blackburn's logic of attitudes, the (purely propositional) deontic operators of obligation and permission are interpreted so as to emphasize an expressive use that reflects the desire to consistently realize specific *goals* or *aims*. On this general approach, Blackburn says that

> [i]f the apparatus of deontic logic can be taken over while [the expressive]
> use is kept primary, so much the better [...] [T]he logical apparatus should
> apply wherever we have the idea of a goal or aim, and corresponding idea
> of something to be avoided, or not to be avoided (Blackburn 1988:509).

This is the first step in the direction of providing a more precise account of inconsistency, in that attitudes and commitments are tied to *aims*, to performing actions associated with the realization of goals. Underlying this is the notion of a *perfect world*, a consistent state of affairs that is only realized if all obligations in the actual world are fulfilled.[5] This notion is used to give an interpretation of the basic non-cognitive attitudes of approval and disapproval, which are modeled by a pair of sentential "Hooray!" and "Boo!" operators $H!$ and $B!$.[6] The distinction between

---

[5]See Hintikka (1969:189-190) for a more thorough discussion of perfect worlds and their origins in Kant's notion of a 'Kingdom of Ends', and see Hale (1993:345) for more details on consistency and model sets.

[6]Recall that in the slow track, $H!$ and $B!$ were applied to descriptions of acts. Here, they are intensional operators, and function just as the deontic operators do.

these is made more fine-grained by supplementing them with a permission operator $T!$ that functions as the dual of $H!$ (so that $H! \equiv_{df} \neg T!\neg$), where

> $H!p$ can be seen as expressing the view that p is to be a goal, to be realized
> in any perfect world. A world in which $\neg p$ is less than ideal, according
> to this commitment. The contrary attitude $B!p$ would rule p out of any
> perfect world, and corresponding to permission we can have $T!p$, which is
> equivalent to not hooraying $\neg p$, that is, not booing p (Blackburn 1988:508).

So $H!p$ expresses an endorsement of p, but gives more structure to the basic attitude of approval, for it is viewed as a goal to be attained in a perfect world, as specifying a *route to an ideal*. Appropriating deontic logic gives more structure to the expressivist semantics, for the attitudes of approval and toleration are modeled by complex relations that are established between the actual world and a set of deontic alternatives approximating a perfect world.

Blackburn makes a pair of crucial moves here. First, he shirks the descriptive talk of deontic logic by interpreting the deontic operators expressively with $H!$ and $T!$, thus modeling the expressive character of moral predicates. Second, he interprets deontic alternatives as *approximations to the ideal*, much like possible worlds that approximate a perfect world where no obligation is left unfulfilled. Roughly, the picture we have here is that of a system of possible worlds that are ordered in logical space by an accessibility relation and the operators $H!$ and $T!$.[7] And this ordering

---

[7] The accessibility relation is a variant of one given by Hintikka (1969:185), and is defined by Blackburn (1988:513-514) in terms of four conditions that instruct how to construct a set of *final ideals* $\{L^*, L^{**}, ...\}$ from an initial set of sentences L:

(Ii) If $H!p \in L$, then $H!p \in L^*$
(Iii) If $H!p \in L$, then $p \in L^*$
(Iiii) If $T!p \in L$, then a set L^* containing p is to be added to the set of next approximations for L
(Iiv) If L^* is a next approximation to the ideal relative to some set of sentences L, then, if $p \in L^*$, $p \in$ subsequent approximations to the ideal L^{**}, L^{***},....

For a criticism of rule (Iiv), see Hale (1993:348-349).

traces out paths or routes from the actual world to an ideal world *via* approximations to the ideal, providing the additional structure required to explain logical relations between moral sentences.

And with this an expressive analogue of logical consistency falls out, in the form of what what Blackburn calls a "consistently realizable attitude," which hinges on the notion of goals that are realized in perfect worlds. Addressing the problem of expressive inconsistency, Blackburn says,

> The solution, I believe, must rely upon a notion of consistently realizable attitude. If you promote p and promote $\neg p$, you are in a practical analogue an inconsistent position, in which not all of your goals can be realized (Blackburn 1992:949).

The notion of inconsistently realizable attitudes is intended to mirror the truth-functional notion of an inconsistent set of sentences, albeit without any overt appeal to truth or truth-conditions. And this sets the stage for a quasi-realist definition of validity in terms of inconsistent realizability, an analogue of the standard definition of validity (in terms of logical inconsistency).

Now let's briefly return to the argument **(F-G)** to see how this additional logical structure might explain the apparent validity of the argument. Again let's start with the unembedded moral sentence (1), which states that lying is wrong, or

(P1′) $B!$(lying)

On Blackburn's account this is *not* a descriptive sentence with ordinary truth-conditions. Rather, its function is to rule lying out of any perfect world. So *no* next approximation to the ideal would allow lying. Then what of (P2) in the Frege-Geach argument, which states that if lying is wrong then getting your brother to lie is wrong?

Well, given that the operator $B!$ is a sentence-forming operator, any sentence to which it is affixed can be embedded into a complex context without violating the Frege point. This gives

(P2′) $B!(\text{lying}) \rightarrow B!(\text{getting your brother to lie})$

which says that *if* lying is ruled out of any perfect world, *then* getting your brother to lie is ruled out of any perfect world.[8] That is, there is no next approximation to the ideal where both lying and not getting your brother to lie are consistently realizable. And this, taken in concert with (P1′), is enough for Blackburn to declare that **(F-G)** is valid.

This is an admittedly rough demonstration of validity, and it obscures the logical structure that licenses each move in the demonstration. Since Blackburn doesn't give a detailed proof, I'll sketch a quick one here, based on the information available in (Blackburn 1988) and (Hintikka 1969). The proof starts with an initial set of sentences $L = \{B!l, B!l \rightarrow B!b, \neg B!b\}$,[9] and then proceeds by showing that this set is unsatisfiable by constructing a tableau where each route to a final ideal contains a formula and its negation. Blackburn's official definition is

> A set of sentences L is unsatisfiable iff each route to a set of final ideals S results in a set of sentences S one of whose members contains both a formula and its negation (Blackburn 1988:514).

And this amounts to showing that the *negation* of

(i) $H![(B!l \wedge (B!l \rightarrow B!b)) \rightarrow B!b]$

---

[8]Compare this with the slow-track analysis, which took the conditional to be an higher-level attitude of approval represented by $H!(|B!(\text{lying})| \Rightarrow |B!(\text{getting your brother to lie})|)$.

[9]Here $l = $ "lying", and $b = $ "getting your brother to lie".

generates a tableau which terminates in inconsistent final ideals.[10] And since $H!$ and $T!$ are duals, that's just to show that we can do the same with the sentence[11]

(ii) $\qquad\qquad T![(H!\neg l \wedge (T!l \vee H!\neg b)) \wedge T!b]$

Now, what (ii) says is that the sentence $H!\neg l \wedge (T!l \vee H!\neg b) \wedge T!b$ is tolerated, or consistent with an ideal world. So there is a next approximation L^* containing $H!\neg l \wedge (T!l \vee H!\neg b) \wedge T!b$.[12] And this is where we can plant the tree:

(iii) $\qquad\qquad\qquad\qquad\qquad H!\neg l \wedge (T!l \vee H!\neg b) \wedge T!b$
$$\mid$$
(iv) $\qquad\qquad\qquad\qquad\qquad\qquad H!\neg l$
(v) $\qquad\qquad\qquad\qquad\qquad\qquad T!b$
(vi) $\qquad\qquad\qquad\qquad\qquad (T!l \vee H!\neg b)$

$$T!l \qquad\qquad H!\neg b$$
$$\{H!\neg l, \neg l, l\}^{**} \qquad\qquad \{H!\neg l, \neg l, b, \neg b\}^{***}$$

This all looks very impressive, but a pleasing aesthetic is no measure of logical success. So a quick explanation is in order. Lines (iv), (v), and (vi) of the tableau follow from the conventional definition of conjunction in model sets, and reflect the position that being committed to (iii) commits one to each of the attitudes contained

---

[10]Note that there is a $H!$ operator with wide scope in (i). The reason for this is due to Hintikka's original definition of *deontic* validity, which is the kind that we're after here. Hintikka discusses the difference between *logical* validity and *deontic* validity at length (1969:192-193), though Blackburn only mentions it briefly (1988:515).

[11]I am working under the assumption that $B!p = H!\neg p$. The reason why I am using this equality (as opposed to $B!p = T!\neg p$) is due to Blackburn's account of $T!$ in (1988:508). This implies that $\neg B!p = T!p = \neg H!\neg p$, and so $B!p = H!\neg p$.
    But this is at odds with a passage later in the article, where Blackburn says that "$\neg H!p$ [...] expresses that [attitude] of opposition: tolerating $\neg p$ or allowing it as consistent with an ideal world" (1988:511). Depending on how this is read, expressing an opposition to p could be interpreted as $B!p$, which would imply that $B!p = \neg H!p$, contrary to the previous passage. So while I have reservations about the correctness of equating $B!p$ with $H!\neg p$, I will reluctantly carry with using this equality.

[12]Formally, this follows from Blackburn's rule (Iiii).

in the complex.[13] The tree then branches at line (vi), splitting into two distinct paths

or *routes*.[14] This models the idea that the agent is *tied to a tree*, to a non-truth-

functional disjunctive commitment where they are tied to *either* tolerating lying *or*

disapproving of getting their brother to lie. And as Blackburn is careful to note,

this is not equivalent to *either* being tied to $T!l$ *or* being tied to $B!b$, but instead

is a commitment to "accepting the one branch should the other prove untenable"

(1988:512). Then each resulting branch leads to a distinct next approximation to the

ideal, and each approximation contains a formula and its negation, showing that the

original set of sentences L is unsatisfiable.[15] And this is supposed to reflect the idea

that one could not commit to the initial set without inducing some sort of incoherence

akin to expressive inconsistency. Again, the argument **(F-G)** is said to be valid.

This seems to provide a nice tight account of the validity of **(F-G)**, and without

any reliance on truth-conditions or realist machinery. Everything is shifted to the

level of attitudes, which are given a rich enough structure that moral sentences can

be embedded into conditionals without running afoul of the Frege point. And mapping

expressivism onto Hintikka's deontic logic explains how logical relations are mediated,

and it provides a non-truth-functional logic for moral discourse.

The novelty of Blackburn's approach lies in the use of a tableau to represent

the process of reasoning from initial sets of (possibly *mixed* moral and descriptive)

statements to a sentence that specifies how an agent should act. The tableau makes

---

[13]Formally, if $p \wedge q \in m$, then $p \in m$ and $p \in m$ for any model set m containing L (1988:513).

[14]Due to the fact that if $p \vee q \in m$, then either $p \in m$ or $p \in m$ for any model set m containing L (1988:513).

[15]The left ideal $\{H!\neg l, \neg l, l\}^{**}$ is constructed in virtue of the commitment to $T!l$ and rule (Iiii). Then from line (iv) and rules (Ii) and (Iii), the ideal contains both $\neg l$ and $H!\neg l$. The right ideal $\{H!\neg p, \neg p, q, \neg q\}^{***}$ is obtained from line (v) and rule (Iii).

no direct appeal to representative or 'realist' truth, but instead follows rules that specify how to manipulate formulae so as to avoid inconsistently realizable goals. Each node of the tableau is derived either from the initial set or from a preceding node in accordance with the rules for model-sets and the sentential operators. And this gives general *commitment conditions* for sentences in the place of truth-conditions. Speaking of the commitments associated with being tied to a tree, Hale says that

> [t]he proposal, as I understand it, is that we can characterize the sense of compounds whose components may be evaluative, not in terms of their truth conditions, but in terms of the inferential commitment involved in endorsing them (Hale 1993:347).

And Unwin makes similar remarks, stating that "the central idea of Blackburn's semantics is just that the meanings of complex formulae are to be explained by their inferential roles" (Unwin 1999:348, and see also 340), by how commitments figure into inferential relations. The idea is that if one accepts the moral commitment $B!(l)$ then they are required, in some specific (but not fully explained) way, to commit to $B!(b)$, and this is what gives the meaning of the conditional. The burden on the expressivist is then to properly explain the meaning of the conditional in terms of its inferential role, which Price cashes out in terms of an 'inferential disposition', where

> [a] sincere utterance of 'if $[B!(l)]$ then $[B!(b)]$' will be said to indicate that a speaker posesses what may be called an 'inferential disposition' - a mental state such that if the speaker were to adopt the mental attitude associated with the utterance '$[B!(l)]$', she would be led to adopt the mental attitude associated with '$[B!(b)]$' (Price 1994:139).

This is a good start towards giving a non-truth-functional account of the conditional, but it pushes the explanation back to an undefined 'leads to' relation between mental

states, which is not much of an improvement over Blackburn's earlier (1984) 'follows from' involvement relation (see device **d3**).

Things are tigtened up a bit by Lenman (2003), who says that if there are specific *norms* that require a commitment to $B!(b)$ given a commitment to $B(l)$, then "we may readily envisage an analogue of the introduction rule for the conditional where a (quasi-)conditional so constructed would equally license an analogue of *modus ponens* – the corresponding elimination rule" (Lenman 2003:35).[16] This brings inference to the fore, suggesting that inferential relations between attitudes might be modeled purely syntactically, without relying on a truth-table analysis of the connectives.

This has led to suggestions that an expressivist semantics should be more in-line with an inferential role (Lenman 2003:36) or conceptual role (Blackburn 1992:951) account of the logical connectives, an idea that appears concretely not only in Blackburn's commitment-tableaux, but also in Gibbard's normative logic (see Gibbard 1990 and §2.2). In Blackburn's case, the tableaux specify how to properly reason with commitments, whether they be beliefs or non-cognitive attitudes. The model-set rules explicitly determine how compound commitments can be introduced and eliminated, and give the conditions under which an agent is warranted in committing to a non-cognitive attitude or factual belief, given the initial set. And an agent is licensed to infer a conclusion from a premise set if they are simultaneously committed to accepting the premises and rejecting the negation of the conclusion. This appears to give a nice non-truth-functional interpretation of expressive validity and inconsistency. And later, in Chapter 3, I'll be returning to this idea.

---

[16]This goes back to comments made by Hare, to the effect that one knows how to use the conditional if one knows how to use *modus ponens*, an idea that is reflected syntactically in the natural deduction rules for the conditional. Here's Price (1994:140) on Hare's idea: "we are in a position to affirm $p \supset q$ if we know that if we are in a position to affirm p then we can go on to affirm q."

2.1.3 Two Problems With the Fast Track

While the fast track may be an improvement over the earlier slow track, it is not free from problems. This section will discuss two of the most difficult problems that Blackburn's logic of attitudes faces. The first problem has to do with properly explaining *mixed* moral and descriptive sentences, where sentences from the moral and descriptive domains are combined to construct compound sentences and inferences. The second problem is narrower, but more severe, and involves a problem with the *negation* of sentences with a dominant attitude operator.

Problem 1: Mixed Sentences

One of the aims of expressivism is to explain how moral discourse can *appear*, on the surface, to be descriptive and fact-stating, and yet still remain committed to a set of logical and semantic constraints that sharply separate it from ordinary descriptive discourse. The difficulties with providing an explanation for this are highlighted in the case of mixed moral and descriptive sentences like

(6)   If lying is wrong, then my mom will get angry at me

The sentence (6) is grammatically well-formed, and one could easily imagine uttering it in ordinary discourse. But sentences of this sort present expressivism with a problem, for they mix sentences from different domains, with different semantic and logical commitments.

Hale (1993:352) characterizes this problem in terms of a dilemma, which the expressivist must respond to by impaling themselves on one of the two following horns. The *first horn* requires that the conditional be interpreted in the standard manner, as a truth-functional connective. The problem is that if this is done, then the only way

to make sense of (6) is to give the antecedent $B!l$ a descriptive reading, and to treat it as expressing a truth-apt proposition, contrary to its intended expressive interpretation. The point is, as Hale (1993:339,352) emphasizes, and that is also made clear from the original Frege-Geach embedding problem, is that it is hard to see how an expressive interpretation of moral sentences can be retained if the logical connectives are interpreted truth-functionally.

The *second horn* of the dilemma requires that the conditional be interpreted non-truth-functionally, and as primarily expressive of an attitude. This is Blackburn's preferred response to the dilemma, where the conditional is cashed out as a disjunctive commitment of being tied to a tree. In Blackburn's logic of attitudes, the conditional sentence (6) would be translated as $B!l \rightarrow m$, which is equivalent to the disjunction[17]

(7) $\neg B!l \vee m$

and this disjunction is explained expressivistically in terms of being tied to a tree. For the purposes of clarity, let's use the equivalence $\neg B! \equiv T!$ (Blackburn 1988:508) to get the equivalent sentence

(8) $T!l \vee m$

Then, according to the tree-tying interpretation, affirming (8) is to express a commitment to a disjunction, to (*either* tolerating l *or* accepting m). And according to Blackburn's (1988:512) interpretation, this is a single disjunctive commitment, and is *not* correctly interpreted as *either* (being committed to tolerating l) *or* (being committed to accepting m). That is, commitments are not distributive: they do not distribute into disjunctions.

What's the point of this distinction between a commitment to a disjunction and a

---

[17]In this particular case, read 'l' as a description of the act of lying, and read 'm' as a description of my mom being angry at me.

disjunction of commitments? According to Hale, it ends up being "fatal to the claim

that $[B!l \rightarrow m]$ can figure into the major premiss for a step of *modus ponens*" (Hale

1993:353), that inferences of the following form, call it **(Mix)**, are instances of *modus*

ponens:

    1. Lying is wrong

    2. If lying is wrong, then my mom is going to get mad at me

∴ 3. My mom is going to get mad at me

The problem, according to Hale, is that the corresponding tableau, which is supposed

to provide the justification for **(Mix)**, in fact involves a distributive disjunction,

contrary to Blackburn's insistence that distribution of commitments be disallowed.

Here's the fragment of the tableau for **(Mix)** that involves the branching commitment:

$$\vdots$$
$$(T!l \vee m)$$

$$T!l \qquad m$$
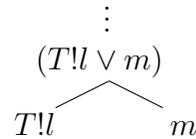
But, says Hale, to go from $T!l \vee m$ to a pair of alternatives $T!l$ and m is "precisely

to treat the commitment as distributive"(Hale 1993:353). The reason that he gives

for this is that it is exactly the same as saying that (in the actual world), *either* you

are committed to $T!l$, *or* you are committed to m.[18] But this is just what Blackburn

wants to deny, and what Hale (1993:353) says must be avoided for expressivism to

---

[18]It is perhaps instructive to compare this with the standard Beth-style semantic tableau, replacing *truth* with *commitment* throughout.

remain credible.[19] So if Hale is right, then if *modus ponens* is to go through, disjunctive commitments *must* be distributive, in opposition to Blackburn's expressive interpretation of them. But doing this results in "a fatal gap open[ing] up between the proposed logic and the preferred interpretation" (Hale 1993:353), exposing a potential flaw in Blackburn's logic of attitudes.

The point is that neither of the horns of the dilemma present palatable options for the expressivist. On the one hand, treating the conditional as truth-functional requires that evaluative subsentences like $B!l$ must be interpreted descriptively, in violation of the Nonfactualism thesis. On the other hand, if compound sentences are treated primarily as expressing attitudes, then expressivism doesn't map properly onto the underlying logical formalism, at least under Blackburn's proposed semantics. Either way, expressivism is in trouble.

Problem 2: Negation

Though Hale's dilemma may have uncovered a flaw in Blackburn's semantics, a more pressing problem is put forward by Unwin (1999), who argues that Blackburn's formalism is unable to account for the simplest embedded context, that of negation. Forget conditionals and *modus ponens*, negation needs an explanation before all else.

Unwin (1999:339) starts by giving the following contextual definition of Blackburn's 'Hooray!' operator (where A is an agent):

$$\textbf{(E)} \quad A \text{ accepts } H!p \equiv \underbrace{A \ \text{ hoorays that } \ p}_{descriptive}$$

The distinction between the descriptive verb "hooray" and the expressive operator $H!$

---

[19]Note that Hale doesn't offer much of an explanation as to *why* he takes this to be the case, though some hints may be found in (Blackburn 1988:513-514).

means that on the l.h.s. A accepts something that is 'marked' by the non-descriptive sentence $H!p$, while the r.h.s. is a descriptive sentence reporting on A's standing in the "hoorays" relation to p. And the equivalence in **(E)** means that accepting $H!p$ is analyzable truth-functionally in terms of hooraying that p. The point of **(E)** is to give a precise definition of the operator $H!$, and hence a more detailed account of the meaning of moral sentences, in terms of the truth-conditions of 'A hoorays that p.'

But Unwin's contextual definition immediately uncovers a wrinkle in Blackburn's semantics. Blackburn introduced a dual $T!$ to $H!$, partially in an attempt to fix the problems with the slow track method, and partially to explain how negated attitudinal operators would be interpreted. One of the virtues of introducing a dual for $H!$ is that it allows for a nice treatment of negation, for it strains the imagination to comprehend what the negation of an *attitude* of approval means, as in the case of $\neg H!p$. But since $\neg H! \neg \equiv T!$, the seemingly problematic $\neg H!p$ is really just the harmless $T! \neg p$ in disguise. And the denial of a derogatory attitude, $\neg B!p$, is just $T!p$. In short, the $H!/T!$ duality allows external negations to be driven inward, so that an attitude operator takes wide scope. So much for the benefits. What's the problem?

The problem is, according to Unwin, that the duality is "just wrong" (1999:341). And the definition **(E)** is supposed to show this, since the following equivalence is implied by it:[20]

(9) A doesn't accept that $H! \neg p \equiv A$ doesn't hooray $\neg p$

The problem with this sentence lies in the l.h.s., which here is obtained by blindly taking the dual $\neg H! \neg p$ of $H!p$, just as it appears in **(E)**. But the result isn't what *toleration* amounts to. According to Unwin, toleration is 'accepting not-$H! \neg p$', the

---

[20]Unwin omits the agent A in his analysis, which complicates things slightly.

acceptance of a negated attitude. The difference here is between accepting the negated attitude not-$H!\neg p$, which is what toleration is supposed to be, and refusing to accept the attitude $H!\neg p$, which is what the definition gives us.

So there appears to be a gap between the concept of toleration and Blackburn's formal representation of it.[21] Blackburn's interpretation of the attitudes of endorsement and toleration in terms of the dual operators $H!$ and $T!$ may be unwarranted, and it may be, as Unwin says, "at best a mere brute fact that the attitudes conflict with each other, with no internal complexity that could explain why" (Unwin 1999:342). While giving a dual to $H!$ is required in order to drive external negations inwards, cashing the resulting dual $T!$ out as *toleration* might be an *ad hoc* move, motivated more by technical expediency than a tight conceptual connection between toleration and endorsement.

So what's the core problem? Unwin diagnoses it as a "syntactic defect" (1999:341) in **(E)** that does not account for all possible placements of the negation operator. There are three possible places to negate the l.h.s. of **(E)**, but only two ways to negate the r.h.s. To see this, look at the following three variations of **(E)**:[22]

(10) *A* does not accept $B!$(lying) $\equiv$ *A* does not disapprove of lying

(11) *A* accepts not-$B!$(lying) $\equiv$???

(12) *A* accepts $B!$(not-lying) $\equiv$ *A* disapproves of not lying

The problem with (11) is that there is no position to place a negation on the r.h.s.

---

[21] Of course this all hinges on the legitimacy of the definition **(E)**. Unwin doesn't show explicitly that it maps onto Blackburn's notion of 'hooraying', merely claiming that "the '$H!$' operator is definable contextually by [it]" (Unwin 1999:339).

[22] This is a variation of an example taken from Unwin (1999:342), though it is modified to accord with the talk of this paper of the wrongness of lying. Strictly speaking, "disapproves" should be read as "boos" here.

without the sentence collapsing into either (10) or (12). The syntactic structure of "A hoorays that p" allows for only two sorts of negation, as displayed in (10) and (12), while the l.h.s. can be negated three ways. The bottom line is that there is a lack of structure that leaves the embedding of $B!$(lying) into a negated context unexplained. This should seem somewhat familiar, since it is really just the Frege-Geach embedding problem, applied to negated contexts.

The bottom line is that there is no tight logical connection between "lying is wrong" and "lying is not wrong" when the predicate is read expressivistically, and this lack of connection can't be fixed by simply wedging opposing or disagreeing attitudes into a rigid logical duality like Blackburn's (mistaken) $H!/\neg T!\neg$. And the problem here cuts deep, since if there is no clear logical conflict between "lying is wrong" and "lying is not wrong," then it becomes difficult to provide a tight account of inconsistency and validity.

The crux of the issue is that it is relatively straightforward to say what it means for *beliefs* to be inconsistent, as they are anchored in states of affairs. But we don't have the same luxury with non-cognitive attitudes, for their direction of fit to the world goes in the opposite direction. What the expressivist needs to do to solve this problem is to give a semantics for negation that gives content to sentence (11), to "S accepts that lying is *not* wrong." Perhaps the most obvious way of doing this would be to find an expressive parallel to descriptive sentences of the form "S believes that snow is *not* white." But again, the problem comes down to negating an attitude, to interpreting a sentence of the form "lying is *not* wrong."

The standard expressivist strategy is to postulate a *new* attitude, distinct from, but somehow related to, the attitude of disapproval. On this approach, (11) becomes

(11^{new}) A accepts not-$B!$(lying) $\equiv$ A (stands in the *new*-attitude
relation to) lying

But once a new attitude is introduced to explain (11), an infinite hierarchy of attitudes is not far behind, for *every* negated context will require a new attitude in order to drive the negation inwards. And if anything violates the parsimonious tendencies displayed by Frege in §1.2.2, it would be an infinite hierarchy of (possibly only loosely related) attitudes. It seems that negation is a problem.

So the criticisms leveled by Unwin and Hale cast a shadow over Blackburn's logic of attitudes, for they uncover several subtle flaws that are indicative of a disconnect between the expressive use of moral language, on the one hand, and the deontic logic that Blackburn deploys to model it, on the other. It appears as though the deontic operators of obligation and permission are not optimally interpreted in terms of the simple attitudes of endorsement and permission. And this stems, in part, from the attempt to interpret a logical formalism that is designed, in the first place to cohere with the realist concepts of truth, satisfaction, and extension.

In the next section I'll look at Allan Gibbard's normative logic, which provides a much richer account of the non-cognitive mental states of approval and disapproval than Blackburn's simple $H!$ and $B!$ operators do. And I'll show that while giving this richer content appears to provide an elegant solution to the embedding problem, it may cause Gibbard's expressivism to collapse into a cognitivist meta-ethical theory. Again, I'll be concerned primarily with the relatively narrow problems of explaining the embedding of (1) into the conditional (P2), and explaining the apparent validity of the argument **(F-G)**.

2.2 Gibbard's Norm Expressivism

Gibbard's (1990) norm-expressivism is *narrowly non-cognitivistic*, in that a sincere utterance of the sentence

(1) Lying is wrong

doesn't say what it is for lying to *be* wrong, doesn't attribute a property *to* the act of lying, and doesn't "state a matter of fact, either truly or falsely" (Gibbard 1990:8). Rather, an utterance of (1) amounts to performing the act of *judging* lying to be wrong, to expressing a non-cognitive state of mind: that of accepting a system of norms that forbids lying.

There's a lot going on here, so let's unpack this a bit. A system of norms N is said to be a weighted set of *descriptive* sentences, and is characterized by four classes of predicates: forbidden, optional, required, and permitted. Then a moral judgment is made by applying a system of norms to a *circumstance*: either to a set of facts that bear on what an individual does, regardless of their epistemic access to those facts (this is an *objective* circumstance), or on what the agent knows or has reason to believe (a *subjective* circumstance). Then an act is *advisable* if applying a system of norms N to a set of objective circumstances returns that the act is N-permitted, and an act is *rational* if applying N to a set of subjective circumstances yields that the act is N-permitted.[23]

This sets the stage for the basic norm-expressivist analysis of unembedded moral claims like (1). For any agent A and alternative S, the norm-expressivist analysis yields the following equivalence, which I see as something akin to a *translation schema*

---

[23]Gibbard's analysis is given in terms of the base property of "rationality," which can be extended to an account for moral predicates. In what follows I'll sometimes conflate "rational" with "right," "irrational" with "wrong," and "normative" with "moral."

linking the normative to the descriptive (Gibbard 1990:91):

$$\textbf{(TS)} \quad A \text{ thinks that } \underbrace{S \text{ is rational}}_{normative} \quad \text{iff} \quad A \text{ thinks that } \underbrace{S \text{ is } N\text{-permitted}}_{factual}$$

where thinking that S is rational is a combination of a normative state and factual belief. More precisely, thinking that an act S is rational amounts to A accepting a system of norms N such that she believes that she is in subjective circumstances for which N permits S.

This nicely ties every normative judgment to an N-corresponding descriptive sentence, but is still fine-grained enough to keep the meaning of "lying is wrong" distinct from "I disapprove of lying," for the former does not express a proposition or have any truth-conditions. Rather, it expresses a non-cognitive mental state that specifies *sincerity conditions* for the sentence. As Gibbard says,

> A normative sentence, the expressivist says, expresses a state of mind; its meaning is explained not by giving its truth conditions but by telling what state of mind it expresses (Gibbard 1990:92).

> When one says one is in a certain state of mind, then actually being in that state of mind constitutes speaking truly. When one expresses a state of mind, on the other hand, being in that state of mind constitutes not speaking truly but being sincere (Gibbard 1990:84).

So what Gibbard's norm-expressivism gives is a non-truth conditional, non-factualist and expressive account of unembedded moral claims like (1). So far so good.

But the current analysis is still susceptible to the Frege point, as lying is not judged to be wrong when (1) is embedded into the antecedent position of a conditional like (P2). Gibbard is well aware of this, and addresses the problem by framing norm-expressivism in a modified possible worlds semantics, a *normative logic*. His aim is to

induce additional structure by leveraging the semantic and logical power of possible worlds, but in such a way that the central expressivist tenets, especially narrow non-cognitivism, are preserved.

The novelty is to replace standard possible worlds with an expressivist-friendly analogue, a *factual-normative* world. This is just a pair of objects $\langle w, N \rangle$, composed of an ordinary possible world w, and a *complete* set of norms N. Since N is complete, any individual who accepts N and applies it to a circumstance in w is able to determine conclusively if a given alternative is forbidden, required, optional or permitted. And if a normative sentence is N-permitted in $\langle w, N \rangle$, then it is said to *hold* in that factual-normative world.

This is quite nice, since it sets up a norm-expressivist approximation of ordinary possible worlds semantics. For just as the content of a descriptive sentence like "I disapprove of lying" can be represented by the set of possible worlds where the sentence is true (relative to context), the *normative content* of the mental state expressed by a sentence like (1) can be represented by the set of *factual-normative* worlds where the statement holds.[24] And finally, the semantic picture can be filled in by capturing meaning with the slogan that "meaning lies in what is ruled out"(Gibbard 1990:99). That is, the meaning of a normative sentence is represented by the complement of its normative content, by the set of factual-normative worlds where the sentence does *not* hold:

> What, then, is a person doing when he voices a complicated normative thought? [...] Cryptically but accurately, I could say [...] that he expresses

---

[24]Formally, for a normative statement S, the normative content O_S of S is represented by the set of fact-norm pairs where S holds, that is $O_S = \{\langle w, n \rangle \mid S$ holds in $\langle w, n \rangle\}$, where S holds in $\langle w, n \rangle$ iff the system n, as applied to the immediate subjective circumstances that the agent finds herself in (in w) permit the action.

> a set of factual-normative worlds. Better, I could say he expresses his
> ruling out certain combinations of factual possibilities with norms. Best,
> though, I can say this: *he expresses a thought that gets its meaning from*
> *its logical ties to other statements, and through them not only to sense*
> *experience, but also to normative governance.* The formalism represents
> these ties (Gibbard 1990:102) (emphasis added).

This is a rich account of the meaning of normative sentences, and it provides a more
complex explanation of the basic non-cognitive attitudes of approval and disapproval
than the simple speech-act analysis of moral sentences given in §1.1. It is even more
structured than Blackburn's (§2.1.2) $H!$ and $T!$ operators, for it imposes the additional
constraint of *normative governance*, which connects normative content to the world
by tying the act of making a normative judgment to a "special kind of motivation"
(Gibbard 1990:100); a motivation to act in accordance with the normative belief
that results from applying a system of norms to a subjective circumstance. Finally,
the inferential relations between pieces of normative content and factual beliefs are
represented by the underlying set theory of the factual-normative world apparatus:

> Propositions and normative content are tied to the world through a web
> of inference. It is here that the formal representation I have given plays
> its part. Given the representation, we can treat entailment in the usual
> way: content P entails content Q if and only if Q holds in all of the
> factual-normative worlds in which P holds (Gibbard 1990:101).

And this provides Gibbard with all of the machinery required to give a basic logic and
semantics for normative sentences. By latching onto the possible worlds framework,
Gibbard is able to blunt the Frege point, as the meaning of a simple unembedded
sentence like (1) is *not* given in terms of a non-assertoric speech act whose meaning
shifts when it is embedded into an unasserted context. Rather, its meaning is given

by the set of factual-normative worlds that represent the non-cognitive state of mind it expresses, by the normative content of the judgment. And then the underlying set-theoretic structure of the factual-normative worlds semantics can be deployed to specify the normative content of a complex conditional like (P2), recursively and in terms of the normative contents expressed by its subsentences: to make a normative judgment by uttering (P2) is to express a piece of normative content that is represented by the set of all factual-normative worlds where "lying is N-forbidden" holds, intersected with the set of all factual-normative worlds where "getting your brother to lie is *not* N-forbidden" holds. So much for the problem of embedding.

Earlier in the paper I spoke of the need to issue proper semantic import permits to regulate embeddings into complex contexts. In Gibbard's system, import permits are not required, for the bridge between the normative and descriptive built out of the translation schema **(TS)** allows for unrestricted travel between asserted and unasserted contexts without fear of reprisal from the truth-conditional semanticist. Every normative sentence N-corresponds to a descriptive sentence *via* the translation, and so every embedded normative sentence can be substituted for a corresponding descriptive sentence that specifies if an alternative is N-forbidden, required, obligated, or permitted. This allows the expressivist to transition between the descriptive and the moral at will, and so the problem of embedding moral sentences into unasserted contexts can be avoided altogether, for the translation schema allows purely descriptive sentences to be embedded in their stead.[25]

And it is this move that gives a nice explanation of the apparent validity of the

---

[25]More formally, if S is a complex sentence containing normative terms or predicates, then S can be translated into a corresponding *descriptive* sentence S_N by systematically replacing each occurrence of a normative predicate in S with its N-corresponding descriptive predicate. And this is tightened up with the logical equivalence (Gibbard 1990:96) **(Hold)**: S holds in $\langle w, N \rangle$ iff S_N holds in w.

argument **(F-G)**, since for any system of norms N we can systematically replace each occurrence of "wrong" in the argument with its N-corresponding descriptive predicate, "N-forbidden." This replacement yields the following descriptive correlate of the moral argument **(F-G)**, call it **(F-G^R)**:

$\quad$ P1^R. Lying is N-forbidden

$\quad$ P2^R. If lying is N-forbidden, then getting your brother to lie is N-forbidden

$\therefore$ C^R. $\quad$ Getting your brother to lie is N-forbidden

Then the set-theoretic operations on factual-normative worlds ensure that the inference is justified: the argument **(F-G)** is said to be valid.[26]

---

[26] The proof is not essential, and can be skipped. The only interesting part lies in the invocation of **(TS)** and **(Hold)**, since if either of these end up being rejected, then the proof is blocked.

$\quad$ *Proof*: Let "forbidden," "obligatory," "required" and "permitted' be represented by N_F, N_O, N_R, and N_P, respectively, where $N_P = N_O \lor N_R$. Since N-forbidden is the negation of N-permitted, we can write $N_F(x) = \neg N_O(x) \land \neg N_R(x)$. Let $l =$ "lying", and let $b =$ "getting your brother to lie". Then we get the following equivalences:

$\quad$ P1^R $\equiv N_F(l)$
$\quad$ P2^R $\equiv N_F(l) \rightarrow N_F(b)$
$\quad$ C^R $\equiv N_F(b)$

Let the standard notation "$w \models N_F(x)$" denote "$N_F(x)$ holds in w". Then by **(TS)** and **(Hold)**, we have the following equalities:

$\quad O_{P1} = \{\langle w, N \rangle \mid$ "lying is wrong" holds in $\langle w, N \rangle \}$
$\quad\quad = \{\langle w, N \rangle \mid w \models N_F(l)\}$
$\quad O_{P2} = \{\langle w, N \rangle \mid$ "if lying is wrong then getting your brother to lie is wrong" holds in $\langle w, N \rangle \}$
$\quad\quad = \{\langle w, N \rangle \mid w \models \neg N_F(l) \lor N_F(b)\}$
$\quad O_C = \{\langle w, N \rangle \mid$ "getting your brother to lie is wrong" holds in $\langle w, N \rangle \}$
$\quad\quad = \{\langle w, N \rangle \mid w \models N_F(b)\}$

Now here is the pudding: all that we need to show is that the proper containment relations between world-norm pairs obtain. That is, we need to show that $(O_{P1} \cap O_{P2}) \subseteq O_C$.

$\quad$ So let $\langle w, N \rangle \in O_{P1} \cap O_{P2}$. Then $\langle w, N \rangle \in O_{P1}$ and $\langle w, N \rangle \in O_{P2}$, so $w \models N_F(l)$ and $w \models \neg N_F(l) \lor N_F(b)$. That is, $w \models N_F(l) \land (\neg N_F(l) \lor N_F(b))$. But this is equivalent to $w \models (N_F(l) \land \neg N_F(l)) \lor (N_F(l) \land N_F(b))$, which is to say that either $w \models N_F(l) \land \neg N_F(l)$ or $w \models N_F(l) \land N_F(b)$. The former case is an absurdity, so we may safely discard it. In the latter case, it follows immediately that $w \models N_F(b)$. But then by the construction of O_C, we have $\langle w, N \rangle \in O_C$.

$\quad$ Therefore $(O_{P1} \cap O_{P2}) \subseteq O_C$, and we're done: **(F-G)** is valid.

What happens here is that the *apparent* validity of the moral argument **(F-G)** is parasitic on the '*actual*' (truth-functional) validity of its N-corresponding argument **(F-G^R)**. And the norm-expressivist's idea behind this take on validity is that if an agent is committed to the premise set, then they express a piece of normative content that entails the content expressed by the conclusion. That is, they could not *coherently* be in the state of mind expressed by the premise set *and* be in the state of mind expressed by the negation of the conclusion. And this is all supported at the level of logic with the set-theoretic operations on factual-normative worlds.

The apparatus seems to get everything down: it explains the problems of logic and embedding, and it gives a compositional expressivist semantics for normative sentences. So it looks like Gibbard might be able to explain the embedding of moral sentences by invoking the logical and semantic structure of a modified possible worlds formalism.

2.2.1   Some Problems With Norm Expressivism

But by using a possible worlds framework, Gibbard runs the risk of diluting the original expressivist position by letting too much realist or cognitivist machinery creep in through the back door. This would be manageable if the central tenets of narrow non-cognitivism were retained, but there is sufficient reason to suspect that Gibbard's normative logic falls short of its anti-realist and non-cognitivist aspirations.

Blackburn (1992) in particular questions if Gibbard has surreptitiously made use of any realist machinery by illegally crossing the "Fregean abyss" (this is a metaphor for the familiar semantic disconnect between the moral and the descriptive incurred by the Nonfactualism thesis – it firmly plants ordinary descriptive judgments, truth-apt

propositions, and truth-functional logic on one side of the abyss, and non-cognitive attitudes and moral judgments on the other).

Blackburn (1992:948) points to two instances where material from the wrong side of the abyss may have been used. *First*, the use of a system of norms may invoke too much realist material, for the systems are composed of descriptive sentences that are applied to circumstances to render normative judgments. The problem lies with what I've called **(TS)**(see p.54), the translation schema that bridges the normative with the descriptive, which is essential for Gibbard's proposed solution to the problems of embedding and logic. The *second* potential problem, which is "hidden deeper in the theory," involves what I've called the **(Hold)** relation (see ft.25), which links the normative relation of 'holding in $\langle w, n \rangle$' with the usual model-theoretic notion of a sentence holding at a world w. Here is what Blackburn says about it:

> [...] a statement with normative terms in it "holds" or not at a factual-normative world $\langle w, n \rangle$, and this is defined in terms of whether a related statement in which the normative terms are replaced by a term that n-corresponds with them, holds at $\langle w \rangle$. The n-corresponding term is purely descriptive, so this last is a classical semantic evaluation (Blackburn 1992:949).

We knew this already, but the problem that was skimmed over in the previous section is that if every normative sentence S is equivalent to an N-corresponding descriptive sentence S_N, and S holds in $\langle w, n \rangle$ iff S_N holds in w, then it's hard to see how the expressivist interpretation of normative content can be retained without making at least *some* commitments to truth-conditions and facts in w, unless satisfaction-in-w can be removed from the analysis. So while using **(TS)** and **(Hold)** was essential to proving the validity of the moral argument form **(F-G)** (see ft.26), accepting these

two conditions may end up *closing* the gap opened up by the Fregean abyss, in the process making norm-expressivism collapse into a cognitivist or realist theory.

These suspicions are confirmed by Dreier (1999), who uses a simple transformation to show that Gibbard's factual-normative worlds $\langle w, N \rangle$ are actually *isomorphic* to functions from sets of norms N to sets of possible worlds (or propositions). More precisely, these functions are like *characters* in a Kaplan semantics (see Kaplan 1989a, 1989b). And under the transformation, the normative content of a sentence like (1) or (P2) can just as easily be represented by a function from sets of norms to propositions:

> Gibbard's device for representing the contents of normative statements turns out to represent them as functions from complete sets of norms to sets of possible worlds. As long as we are happy to use sets of possible worlds as propositions, we can think of Gibbard's device as representing the contents of normative statements by functions from sets of norms to propositions (Dreier 1999:561).

This is not an entirely surprising result, since normative content, or the semantic value of a normative sentence, is determined by the *expression* by an agent A of a system of norms N that is applied to a circumstance to yield a normative judgment.[27] That is, there are explicit contextual parameters that determine the semantic value of a normative judgment, namely the agent A and the normative system N. So in much the same way that the extension of an indexical sentence like "I disapprove of lying" is determined by certain salient features of a context of utterance, the semantic value of (1) is determined by, or *is represented by*, its character, a map from contexts to propositions. Normative systems are just contextual parameters in an ordinary

---

[27]A minor point to note is that Gibbard's notion of "circumstance" is different than Kaplan's "circumstance of evaluation."

Kaplan-style semantics, and this semantic framework applies to *all* of language, descriptive and moral inclusive. As Wedgwood says, referring to Gibbard's normative logic,

> [i]n effect, it is a semantics for the entire language, consisting of a variant of possible-worlds semantics, with an additional parameter that makes a difference only the case of statements involving normative terms (Wedgwood 1997:82).

And while this result gives us some insight into the structure of Gibbard's system, it cheapens his overall result, since it becomes clear that his purportedly expressivist and narrowly non-cognitivist theory is actually just a redressing of a Kaplan-style semantics, complete with extensions, propositions and truth-conditions.

There are two important results that I think should be taken away from Dreier's transformation, one negative and the other positive. Both of these results stem from an observation that Dreier makes, a "payoff," as he calls it;

> And now for the payoff: insofar as we prefer to have some other sort of abstract object to play the role of propositions in our semantic theory, we can just take the contents of normative statements to be functions from sets of norms to those objects (1999:561)

The *negative* result follows from a move that Dreier makes after stating the payoff, which is to take these objects to be "ordinary factualist" propositions (1999:563). This is initially treated as a *bonus* by Dreier, as it explains our commonsense attributions of truth to moral claims. So the quasi-realist gains a point with this move.

But this may not be such a bonus, since invoking factualist propositions deprives

expressivism of much of its anti-realist clout, violating the basic tenets of narrow non-cognitivism: an utterance of "lying is wrong" in a context expresses a truth-apt proposition, in much the same way that "I disapprove of lying" does.[28] In fact, it's hard to see a substantial semantic difference between the two sentences. And it's important to note that this isn't a feature that's unique to the *transformed* norm-expressivist apparatus, since that apparatus is *isomorphic* to Gibbard's original theory. Narrow non-cognitivism appears to be in trouble.

But Dreier's payoff can also be cashed in to yield a *positive* result. For just as the expressivist can map norms to factualist propositions, they are free to map to any other appropriate abstract object (cf. Dreier 1999:561) that plays the role of a proposition in the theory:

> But suppose we had some other sort of proposition in mind. We could borrow the compositional semantics for that sort of proposition, and the account of logical implication too [...] The general solution is to see that once Gibbard's sets are transformed into functions from norms to ordinary propositions, any kind of proposition works equally well (Dreier 1999:562)

So why not map to an expressivist-friendly proposition, a non-factualist proposition that does not represent states of affairs in the world, and is not assigned representative truth-values? And then why not define a logical consequence relation that preserves a non-factualist semantic concept in the place of truth? And then, finally, why not try to give a meaning theory with these notions in mind? The next chapter will be concerned with answering these questions.

---

[28]Where "I disapprove of lying" is just shorthand for the long-winded "I accept a system of norms N which, when applied to my exact subjective circumstances, N-forbids the act of lying."

Chapter 3

Expressivism and Truth

Gibbard and Blackburn's programs have shown how pre-existing deontic and possible worlds frameworks can be modified to give a compositional expressivist semantics. The trick with their approach is to strike a delicate balance between properly formalizing non-cognitive attitudes, on the one hand, and avoiding an overcommitment to moral facts, truth-conditions, and propositions, on the other. Both Gibbard and Blackburn attempt to walk this fine line by making subtle modifications to a model-theoretic framework, assigning moral sentences like (1) to non-cognitive mental states, and then giving a recursive compositional semantics in terms of operations on these states.

However, the problems identified in the previous chapter suggest that neither program succeeds in striking this delicate balance, either providing an inadequate formal model (Blackburn) or deploying too much realist machinery (Gibbard). Blackburn's logic of attitudes goes awry in that the deontic operators of obligation and permission are not optimally interpreted in terms of the simple attitudes of endorsement and permission. And while Gibbard's factual-normative worlds are designed to model

normative content without deviating from narrow non-cognitivism, Dreier's transformation shows that norm-expressivism is in danger of collapsing into a cognitivist or realist theory.[1]

This chapter will delve a bit more deeply into the nature of truth and the role that truth and truth-conditions play in sophisticated expressivism. The chapter is structured as follows: section §3.1.1 investigates deflationary truth, and argues that while moral sentences have deflated truth-conditions, they require a more substantial notion of truth-condition if they want to solve the "deepest" problem of embedding. Then in §3.1.2, alethic pluralism will be investigated, and in §3.1.3 two problems with it will be given. The first problem involves mixed moral and descriptive sentences and inferences. The second problem concerns a dilemma due to Wright (1988). In §3.2 the direction of the chapter shifts slightly, and focuses on potential replies to this dilemma. Section §3.2.1 finds common ground between Hare (1970) and Blackburn (1988) with their inference-rule definitions of the conditional. Then in §3.2.2 some brief comments will be made about proof-theoretic semantics, and how it relates to expressivism. Finally, in §3.3.3 we return to Dummett's (1973) discussion of the Frege point, and ask if his notion of a *quasi*-assertion might help blunt the second horn of Wright's dilemma.

---

[1]It should also be noted that Gibbard's norm-expressivism confronts the same problem with negation that Blackburn's logic of attitudes does, though this particular problem was not discussed in §2.2.1. See Unwin (2001) for the details.

3.1 Truth and Truth-Conditions

While clause (i) of the Nonfactualism thesis denies that "lying is wrong" is truth-apt or has truth-conditions, the sentence is still embeddable into simple "It is true that.." and "It is false that..." contexts, for

(13)   It is true that lying is wrong

is grammatically well-formed and plays a significant role in moral discourse. One of the goals of expressivism in general, and quasi-realism in particular, is to account for this phenomenon, and to explain how moral sentences can be denied truth (as correspondence) conditions, and yet still bear all of the overt trappings of ordinary assertoric discourse, including the application of a truth predicate.

3.1.1 Inflated and Deflated Truth

Stoljar (1993) suggests that an explanation may be forthcoming if a distinction is made between *inflated* truth-conditions, which are taken to establish a connection between a sentence and some language-independent fact, and *deflated* truth-conditions, which a sentence trivially possesses if it can be embedded grammatically into "It is true that..." or "It is false that..." contexts (Stoljar 1993:83). On this view, the deflated truth predicate functions as little more than a grammatical device that expresses endorsement when it is applied to a sentence, and so embedding "lying is wrong" into an "It is true that..." context, yielding (13), does not violate the spirit of the Nonfactualism thesis. On the contrary, it suggests that clause (i) of the thesis should be refined to accommodate the distinction between inflated and deflated truth-conditions,

and should instead be read as:[2]

> Nonfactualism:
>
> (i′) Moral sentences are not truth-apt and they do not have truth-conditions, in the *inflated* sense of truth. But they *are* truth-apt, and they *do* have truth-conditions, in the *deflated* sense of truth.

And with this minor modification the apparent truth-aptness of "lying is wrong" is afforded an explanation: the truth predicate applied to "lying is wrong" in (13) is simply *deflated*, and establishes no correspondence between "wrong" and any suspicious properties in the world.

This result is leveraged by Stoljar (1993:91) to give a quick and tidy solution to the Frege-Geach embedding problem. In effect, the proposed solution goes like this: we know that the moral sentence (1) is conventionally used to express an attitude of disapproval, and we know that this attitude is suppressed when the sentence is embedded into the conditional (P2). But (1) is also assigned (deflated) truth-conditions, which are *not* affected by the embedding, and that follow (1) into the unasserted context without being altered. There are two immediate consequences of this. *First*, the (deflated) truth-conditions of the conditional (P2) are determined recursively in terms of the (deflated) truth-conditions of its subsentences. *Second*, the apparent validity of the argument **(F-G)** can be given a simple explanation by taking logical consequence to be necessary preservation of *deflated* truth.

---

[2]This is an amalgam of Stoljar's clause (SP-infl) (1993:84) and the fact that moral sentences have deflationary truth-conditions. Note that I have omitted mention of propositions in this revised version of Nonfactualism.

Recall that in (§1.2.3) it was suggested that the embedding problem might be solved by identifying something essential to the meaning of (1), distinct from expressive force, which 'tagged along' with the sentence into unasserted contexts. What Stoljar is proposing here is not dissimilar to this idea, for it locates the meaning of (1), or at least the meaning that is relevant to the validity of the argument, in the (deflated) truth-conditions of the sentence. This move hinges on a distinction that he draws between the *expressive* or *pragmatic* meaning of a moral sentence (which is given by the attitude that the sentence is used to express) and the (deflated) truth-conditions that are assigned to the sentence.[3] The key point of this distinction is that a difference in the *expressive* meaning of (1), induced by an embedding, does not necessarily entail a difference in the truth-conditions of (1), where it is the truth-conditions of the sentence (and not its *expressive* meaning) that determine if the argument **(F-G)** is valid or not. That is, charges Stoljar (1993:93), Geach's original argument mistakenly conflates the *pragmatic* features of the premises of **(F-G)** with their *semantic* features when it is (incorrectly) claimed that an equivocation in the *expressive* meaning of (1) entails an equivocation in its *meaning* – where it is *meaning* (or truth-conditions) that affects the validity of the argument. So rather than going to great lengths to explain the validity of the argument in the face of an equivocation of expressive meaning, all that is required to explain validity is to demonstrate that there is no equivocation in the (deflated) truth-conditions of its constituents.

At first glance, Stoljar's proposal appears to accomplish in a few short steps what took Gibbard and Blackburn significantly more work to do. By assigning deflated

---

[3]Stoljar uses "emotivism" in the place of "expressivism" in his paper, but there is nothing significant behind this difference in terminology. What he calls emotivism is essentially just the conjunction of clause (i) of the Nonfactualism thesis with clause (i) of the Expression Thesis.

truth-conditions to moral sentences, and by distinguishing those truth-conditions from expressive meaning, Stoljar is able to explain how (1) can be embedded into the conditional (P2), and how the argument **(F-G)** is valid, and all without violating the anti-realist constraints imposed by expressivism.

But there is a problem with this approach, for deflationary truth-conditions are so 'thin' that they have no significant explanatory power. As Lynch says,

> The deflationist picture is extremely attractive. But is comes at a price. That price is that it removes truth from our explanatory toolkit. If the deflationist is right, we are barred from appealing to truth and its nature to help explain other items of philosophical interest: meaning, content determination, knowledge, the norms of belief, and so on (Lynch 2008:123).

This point is driven home by Dreier (1996) and Sinnott-Armstrong (2000), who demonstrate, quite convincingly, that Stoljar's deflationism does not solve the "deepest" problem of embedding: explaining what the meaning of unembedded moral sentences like "lying is wrong" is.[4] If truth-conditions confer a sentence with its meaning, and the truth-conditions of a sentence like "lying is wrong" are deflated, then the deflationist simply can't tell us what (1) means.[5] The upshot is that saying *only* that "wrong" is used as a speech act to express an attitude of disapproval, *and* that "lying is wrong" has deflated truth-conditions, is *not* enough to explain what (1) or the conditional (P2) actually mean. Deflated truth-conditions simply won't cut it for expressivism – something more substantial is required.

---

[4]The original argument is due to Dreier (1996:43), and aimed to show (i) that treating "wrong" merely as a logical predicate is problematic, and (ii) why the more sophisticated accounts of Blackburn and Gibbard are required if an inference-rule account of the conditional is to be given. Sinnott-Armstrong (2000:689) applies Dreier's argument directly against Stoljar's deflationism (which he calls "minimalism").

[5]Again, it is crucial to not conflate the *pragmatic* or *expressive* meaning of (1), with its meaning (or truth-conditions).

This conclusion is reinforced by Sinnott-Armstrong (2000:691), who argues that even the rich semantic content offered by Gibbard's factual-normative worlds isn't enough to close the explanatory gap left open by deflationism.[6] The additional logical structure does little to explain the meaning of moral sentences if it isn't supplemented with a theory of truth that is more substantial than the deflated. Referring again to the "deepest" problem of embedding, of explaining the meaning of unembedded atomic moral sentences, Sinnott-Armstrong says that

> [...] it is hard to see how to solve the deepest problem of embedding without giving truth conditions along the lines proposed by the moral realists. Moreover, it seems natural to fill the gap in expressivism by supplying realist truth conditions for evaluative judgments, and such truth conditions do seem to solve the deepest problem of embedding [...] So expressivists still have a lot of serious work to do. I do not claim that expressivists cannot accomplish this task, but I would like to see them try (Sinnott-Armstrong 2000:693).

So expressivism is committed, at a minimum, to deflated truth-conditions, given that (13) is a grammatical sentence. But the arguments presented by Dreier and Sinnott-Armstrong indicate that, in order to explain the meaning of "lying is wrong," expressivism *must* be committed to truth-conditions that are more substantial than the deflationary. And at the other end of the spectrum, the Nonfactualism thesis, especially clause (ii), places a strict upper bound on the strength of a moral truth predicate, denying that it can be defined in terms of corresponding to moral facts or properties. So how can the expressivist get all of the explanatory power of realist truth-conditions without violating the basic tenets of expressivism?

---

[6]This involves an extension of Dreier's argument that is run on Gibbard's normative logic. For the details, see (Sinnott-Armstrong 2000:688-691). A similar argument is run on Blackburn's logic of attitudes (Sinnott-Armstrong 2000:691-692), and yields similar results.

3.1.2 Alethic Pluralism

At first glance, it looks like an easy way out may be offered by Wright (1988, 1992, 1995), whose *alethic pluralism* postulates that a range of different truth predicates, varying in degrees of metaphysical commitment, apply to different domains of discourse. Though not referring directly to Wright, Tennant captures the essence of the alethic pluralist project when he says

> [t]ruth comes in different kinds. Only one of these – should it be philosophically coherent – suffices for [realism in truth-value]. The other kinds of truth form a sort of crescendo leading up to it. It starts *pianissimo*, in complete philosophical neutrality, with a deflationary conception of truth, and ends with a bang, with full-blown realist truth (Tennant 1997:15).

With a range of truth predicates at our disposal, the nature of the debate between moral realism and anti-realism can be cashed out a bit differently. For the debate can be (and *is*, by Wright) framed entirely in terms of the *kind* of truth predicate that is properly applicable to the sentences of the moral fragment of natural language. On this view, the anti-realist need only construct a 'lightweight' truth that is, in some relevant sense, *stronger* than deflationary truth, and that yet lacks the metaphysical punch of the 'heavyweight' or inflated truth predicate characteristic of descriptive discourse. Says Wright,

> Here is one proposal. We need to win through to a conception of truth which allows us to grant truth-aptitude, and indeed truth, to responsible judgements within a given discourse without thereby conceding a realist view of it. Such a view will hold that to ascribe truth to a statement need not to be to ascribe a property of intrinsic metaphysical *gravitas* [...] (Wright 1995:214)

The balance to be achieved by the moral anti-realist is delicate, and the specifics that Wright offers are complex. In short, he suggests that moral truth be though of in terms of *superassertibility*, where a statement is superassertible if "it is assertible in some state of information and then remains so no matter how that state of information is enlarged upon or improved"(1995:217-218). More intuitively, a moral sentence is morally true if it can be *morally justified*, and if that justification cannot be defeated, no matter how much "additional consideration is given to the matter" (Wright 1995:218).[7] Details aside, Wright's superassertibility is designed to give the moral anti-realist some nonfactualist analogues of the realist notions of truth, truth-condition, assertion and proposition. And it reduces the opposition between moral and descriptive discourse to a distinction between the *kind* of truth that is applicable to sentences in each domain: truth as correspondence or representation for the descriptive, and truth as superassertibility for the moral.

3.1.3 Two Problems With Alethic Pluralism

This approach seems promising, at least initially, for it suggests that the expressivist might be able to chart a path between deflated and realist truth by adopting a suitably irrealist kind of truth, and from that conception of truth constructing corresponding notions of proposition and logical consequence. However there are two unfortunate

---

[7] Lynch gives a more detailed version of superassertibility in his (2008:124-125), cashing moral truth out as *superwarrant*. The interesting aspect of superwarrant is that its underlying logic is intuitionistic, since at any given information state there may be propositions for which there is neither warrant for them nor for their negation (see (2008:133-4)). And superwarrant is given a *constructivist* account of validity, where an argument is valid iff at every possible world where the premises are superwarranted, then so is the conclusion.

problems with this approach. The *first* is one that is purely internal to alethic pluralism, and has to do with mixed sentences and mixed inferences. The *second* problem involves a dilemma, due to Wright, that any expressivist who aspires to take advantage of alethic pluralism must contend with.

Problem 1. Mixed Sentences and Mixed Inferences

Given a truth predicate "$true_M$" for moral discourse, the alethic pluralist can explain the validity of the moral argument **(F-G)** simply by defining a moral consequence relation in terms of necessary preservation of $truth_M$. And the same can be said for descriptive discourse, and its associated truth predicate "$true_D$." But a wrinkle is introduced when sentences from the two domains, which have different truth predicates and (possibly) different logical consequence relations, are combined to generate complex sentences and inferences, as in the case of the familiar mixed moral and descriptive inference **(Mix)**:

> 1. Lying is wrong
> 2. If lying is wrong, then my mom is going to get mad at me
> ∴ 3. My mom is going to get mad at me

In her (1997), Tappolet presents a very brief, but persuasive, argument to the effect that **(Mix)** cannot be valid if there isn't a single *unique* truth predicate that is preserved in the inference.[8] As she says,

> Mixed inferences remind us of a central platitude about truth, namely that

---

[8]Her example mixes the descriptive and the comic domains of discourse, attributing a heavyweight truth predicate to the former and a lightweight truth predicate to the latter. Her example is "Wet cats are funny; This cat is wet; ergo this cat is funny" (1997:209).

> truth is what is preserved in valid inferences. Moreover, they show that all
> sentences which can appear in such inferences are assessable in terms of
> the same truth predicate. The upshot is that only a truth predicate shared
> by all sentences which can appear in inferences will satisfy the platitude
> relating truth to inferences (Tappolet 1997:210).

The problem faced by the pluralist can then be characterized succinctly in the form of a trilemma:[9] they must either (i) deny that **(Mix)** is valid, contrary to the appearance that it is, (ii) come up with a non-standard account of validity, distinct from necessary truth-preservation, or (iii) commit to a unique truth predicate that is applicable to every sentence in the inference.

In response to Tappolet's paper, Beall (2000:381) suggests that the problem can be solved by adopting a three-valued logic, saying that since alethic pluralists "are so-called because they recognize different truth predicates, the lead of many-valued logic is a natural one for them to follow"(Beall 2000:382). In the specific case of **(Mix)**, the alethic pluralist need only assign the numerical values of, say, 1, 1/2 and 0 to true_D, true_M and "neither true_D nor true_M," respectively, and then stipulate that 1 and 1/2 are *designated values*, or as Beall says, "the different ways of being true, as it were" (2000:382). Then, the validity of **(Mix)** can be given a quick explanation in terms of the necessary preservation of designated values. It looks that by adopting a many-valued logic, a quick and tidy solution to the problem of mixed inferences can be given.[10]

---

[9]In his (2000:381), Beall acknowledges that there is a dilemma, but it is not until Tappolet's (2000) that the third horn is explicitly stated.

[10]A complication arises when the truth predicate for the moral domain is taken to be Lynch's superwarrant (see ft.7), for it is based on an intuitionistic logic. It is not immediately clear if Beall's proposed solution would accommodate cases where intuitionistic and classical logic would have to be combined, in addition to truth predicates being mixed. For more on this issue, minus a discussion of many-valued logics, see Lynch (2008:137-139).

But in her (2000) reply to Beall, Tappolet counters that by adopting a three-valued logic, Beall implicitly commits himself to the third horn of the trilemma, for it is "difficult to believe," she says, that "[designated value] is not a kind of truth" (2000:384). She argues that *designated value* plays the role of a general truth predicate, for it is applicable to any sentence that is either true$_M$ or true$_D$. And most importantly, she takes designated value to be the predicate that "does the inferential job" (2000:384) in **(Mix)**. But, Tappolet asks, if there is one general truth predicate that does all of the heavy lifting in inferences, "why should we need the many truth predicates instead of the [general one?]" (2000:384). The lingering question is if alethic pluralism can remain coherent if for every two domain-specific truth predicates it commits to, it necessarily entails a further single, unique and universally applicable truth predicate, in the form of designated value.

The problem of mixed inferences presents alethic pluralism with a sharp challenge, and extends beyond inferences to the more fundamental, but no easier, problems of mixed conjunctions (Tappolet 2000, Edwards 2008) and combining logics (Lynch 2008).

Problem 2. Truth, Assertion and Expressivism

Even if the problem of mixed inferences were to be given a decisive solution, the expressivist who adopts alethic pluralism is bound to encounter a dilemma of an altogether different sort. The dilemma, due to Wright (1988), stems from what he calls the *goal* of quasi-realism (or expressivism in general): "to explain how *all* the features of [moral] discourse that might inspire a realist construal of it can be harmonized with objectivism" (Wright 1988:35). Given this goal, Wright claims that the quasi-realist

is confronted with two horns:

Horn 1: Quasi-realism *fails.* In this case, the surface syntax of moral discourse is left unexplained, and there is no compelling reason for how we can speak *as if* moral sentences are true and false, at least while still retaining some version of the Nonfactualism thesis.

Horn 2: Quasi-realism *succeeds.* In this case, quasi-realism is able to provide a plausible account of ascriptions of truth to moral sentences. The problem with this horn, according to Wright, is that

> then we shall wind up – running the connection between truth and assertion in the opposite direction – with a rehabilitation of the notion that such statements rank as genuine assertions, with truth-conditions, after all [...] in which case it makes good all the things which the projectivist started out wanting to deny: that the discourse in question is genuinely assertoric, aimed at truth, and so on (Wright Quasi:35).

The rub is that if the quasi-realist program succeeds, and a suitably irrealist truth predicate is constructed, then the expressivist is in effect committed to ordinary assertions that affirm the truth$_M$ of propositions. And as Hale remarks, as a result the entire project of treating moral judgments as primarily *expressive* is "at best an unnecessary detour," and at worst a "blind alley"(Hale 1993:351). For by 'winning through' to a concept of moral truth, the crucial semantic distinction between the purportedly expressive sentence "lying is wrong" and the genuinely descriptive sentence "I disapprove of lying" may end up collapsing.

3.2 A Proposed Solution to the Dilemma

Hale (1993:349-351) suggests that to avoid the dilemma and retain the expressive interpretation of moral sentences, the expressivist must first properly disentangle two projects: the *first* is that of constructing a notion of moral truth, of winning through to an irrealist truth predicate that can be applied to moral sentences. The *second* project is that of developing suitably precise notions of consistency and validity for non-cognitive attitudes, as the sophisticated expressivist project aims to do. The general idea is to "effect a clean break" between constructing a logic for inferences containing evaluative components, on the one hand, and "the project of constructing a notion of truth applicable to moral judgments" (Hale 1993:351) on the other. And the aim of properly distinguishing logic from truth would be to relegate the latter concept to the periphery, and to bring logic to the fore by focusing on the role of inference in moral discourse.

In this section I'll outline how this might proceed, first by revisiting the inference-rule explanation of the conditional that figured so prominently in Blackburn's logic of attitudes, and that can be traced back to Hare. Then in §3.2.2, I'll ask how the inference-rule approach can be severed from truth, in accordance with Hale's suggestion. This will be based on remarks made by Blackburn that suggest a *proof-theoretic* semantics for moral discourse.

3.2.1 Inference-Rule Explanations of the Conditional

Unsurprisingly, Hale specifically points to the commitment semantics of Blackburn's logic of attitudes as a potential starting point. The reason is that Blackburn's preferred definition of the conditional, given in terms of non-truth-functional branching commitments, is in essence an inference-rule definition that makes no direct appeal to truth-conditions or a truth-table analysis of the logical connectives. On this point, at least, Blackburn aligns himself with a larger, but I think under-appreciated, trend in expressivism that aims to explain the meaning of the logical connectives in terms of their inferential role, and not in terms of some irrealist truth-surrogate. This trend runs through Blackburn (1984, 1988), Gibbard (1990), and even Horwich (1993),[11] but originates in Hare's (1970).

According to Hare, to know the meaning of a conditional sentence we need only know the following two items (call these the *inference-rule* conditions):

ir1 The meaning of the general conditional form $\phi \supset \psi$, which we know if we know how to do *modus ponens*.

ir2 The meanings of the individual embedded sentences "encaged" in the conditional. And we know these meanings if we know:

    (i) That they are used to make assertions when unembedded.

    (ii) What assertions they are used to make, when they are unembedded.

The first condition can be read purely syntactically, stating that if we know the $\supset$-elimination rule, then we know the corresponding $\supset$-introduction rule. The second

---

[11]Horwich's (1993) proposal is part of a criticism of Gibbard's norm-expressivism, and purports to show that the embedding problem can be avoided if "our entire use of the term ["wrong"] can be explained by" the two independent *rules of use*: (i) "lying is wrong" expresses an attitude, and (ii) "wrong functions logically as a predicate – that is, "lying is wrong" functions logically in accordance with the natural deduction introduction and elimination rules (Horwich 1993:75).

condition is similar in spirit to both Searle's loophole (§1.2.3), and to Blackburn's definition of the conditional in his (1984), for sub-clause (ii) makes reference to the speech-act that an embedded sentence uses when it appears in unasserted context.[12] The conditions make it possible to "transfer" an unembedded sentence into an embedded context *without* having the meaning of the sentence shift.

So what's the point of this? Well, Hare uses it to show two things: first, that the inference-rule conditions provide a plausible explanation for the meaning of the conditional when ordinary *descriptive* sentences are embedded. And with this base case in place, Hare then goes on to show that this same result holds when *moral* sentences are embedded into a conditional. That is, the inference-rule explanation of the conditional explains the embedding of *both* descriptive and moral sentences – it doesn't matter what *kind* of sentence is 'encaged' in the antecedent and consequent positions of the conditional, since the inference-rule conditions do all of the heavy logical lifting.

Here's how the argument runs. First, consider the descriptive conditional

(14)   If the cat is on the mat then it is purring

According to the inference-rule conditions, to know the meaning of (14) we only need to know the meaning of the general conditional form and the assertions that are made by the 'encaged' antecedent and consequent when they are 'not encaged,' or when they are in unasserted contexts. But since we know these things, we know the meaning of (14). As Hare says,

We know, that is to say, what it is to say that if we are in a position to

---

[12]Perhaps a bit more precisely, it is parallel to the idea that the attitude of approval expressed by the sentence $H!\,(|B!\,(\text{lying})| \Rightarrow |B!\,(\text{getting your brother to lie})|)$ is determined recursively by reference to the attitude that $B!(\text{lying})$ *would* express when unembedded, which is referred to by the embedded antecedent $|B!\,(\text{lying})|$ (see (§2.1.1), esp. p.35.).

affirm the categorical "The cat is on the mat," we can go on to affirm the categorical "It is purring" (Hare 1970:17-18).

So far so good, but the case of ordinary embedded assertions is not overly interesting. However, with a minor modification to the inference-rule conditions, this account of the conditional can be generalized to cover a wider range of non-assertoric speech-acts, including expressions of an attitude.[13] Consider the mixed moral and descriptive sentence

(6) If lying is wrong then my mom will be mad at me

Can we say that we know the meaning of this sentence? We know how to use *modus ponens*, and we know the meanings of the 'encaged' subsentences, since in asserted contexts the antecedent is used to express a disapproval of lying, and the consequent is used to affirm the truth of a proposition. So, according to Hare's analysis, we know

(kn) what it is to say that *if* we are in a position to express an attitude of disapproval towards lying, *then* we can *go on* to affirm the truth of the proposition expressed by "my mother will be mad."

So much for the problem of embedding. And since the meaning of the conditional is given a tidy explanation, the apparent validity of the mixed moral and descriptive argument **(Mix)** can also be explained, and without appealing to truth-conditions or the necessary preservation of truth. In effect, *if* one is prepared to say that lying is wrong, *and* they are prepared to say if lying is wrong then their mom will be mad at them, *then* they can *go on* to say that their mom will be mad at them. As Hare

---

[13]The precise modifications would be to the sub-clauses of the inference-rule condition **(ir2)**, which in revised form would read: (i) that they are used to make a specific (assertoric or non-assertoric) speech-act when unembedded, and (ii) what specific (assertoric or non-assertoric) speech-act they are used to make, when they are unembedded. Note that Hare makes no specific mention of these modifications.

puts it, the only difference between the mixed moral and descriptive case, on the one hand, and the purely descriptive case of sentence (14), on the other, is that in the former case

> to affirm ["lying is wrong"] is here to [disapprove of lying]. But this does not make the meaning of ["lying is wrong"] in the categorical premise different from that of the same words in the conditional clause of the hypothetical premise in any sense that is damaging to the inference, any more than the fact that "The cat is on the mat" (categorical) is used to assert that the cat is on the mat, whereas the same words occurring in a conditional clause are not used to make this assertion, invalidated the inference we discussed earlier (Hare 1970:19)

The idea is that since we know how to use the conditional form, and since we know the meaning of the conditional (6) (from **(kn)**), we also know that *if* we affirm that lying is wrong, then we can *go on* to 'let the consequent of (6) out of its cage,' and affirm that the speakers' mom will be mad at them. That is, **(Mix)** is valid.

Hare's account of embedding is rather elegant, as it simply relies on the inference-rule for the conditional to explain embedding and validity. In short, you know how to use the conditional if (i) you know how to use *modus ponens*, and (ii) you know the meanings of the 'encaged' antecedent and consequent. And as Dreier (1996) notes, Hare's inference-rule approach is similar in theme and detail to Blackburn's (1988) tree-tying account of the conditional. Dreier says that while Blackburn's inference-rule explanation of the conditional is similar to Hare's, still

> [a]dmittedly the notion of being "tied to a tree" of a tableau deduction [...] is not spectacularly clear. [Blackburn] might better have said that someone understands disjunction, whether of factual propositions or of

> attitude-expressing judgments, when she understands disjunctive syllo-
> gism, and that she understands the material conditional when she under-
> stands *modus ponens* (Dreier 1996:41).

The direction that this is all pointing to is an inferential role account of the logical connectives, where understanding how to use the introduction rule for a connective follows from knowing how to use the corresponding elimination rule. And on this picture the meaning of an expression is given by its logical role, and not by truth-conditions or reference to facts in the world.

Let's return briefly to Hale's proposed solution to Wright's dilemma (§3.2). There it was stated that to avoid the dilemma, the expressivist was required to "effect a clean break" between the logic of attitudes and the project of constructing moral truth. The inference-rule account of the conditional surveyed in this section suggests a plausible starting point for a non-truth-functional logic of attitudes. But still truth has not been pushed to the periphery, as Hale's proposal requires. What is needed, and what is still missing, is some key semantic concept to play the role of truth in the semantic theory. And to avoid the dilemma, this key concept must not engender an alethic pluralism or any realist commitments, but it still must be structured enough to avoid the problems encountered by Stoljar's deflationary truth in §3.1.1.

In addition, any formal semantic framework deployed must not succumb to the same fate as Blackburn's deontic logic, which lacked the structure to provide a complete account of the basic attitudes of approval and disapproval. The same goes for the factual-normative worlds apparatus of Gibbard's normative logic, which by design is apt to collapse in an ordinary possible worlds semantics, complete with extensions, denotations, and model-theoretic satisfaction. And this points to a systemic problem with the sophisticated expressivist positions surveyed in Chapter 2, namely that

the model-theoretic frameworks deployed by Blackburn and Gibbard are designed, in the first place, to accommodate realism in ontology and truth-value by establishing relations between linguistic items and formal structures representing the world. This suggests that a proper anti-realist and inferentialist semantics should eschew model-theoretic semantics, and opt for a framework that is more conducive to its inferentialist and non-denotational aspirations. As Shapiro says (originally in the context of mathematical anti-realism);

> [t]o put it baldly, if the variables of [moral language] are not to be understood as ranging over a domain of discourse, then they must be understood in some other way [...] This yields antirealism in ontology. And if the bulk of the justified assertions of [moral language] are not true, then some other account of which assertions are warranted or legitimate is requisite [...] This is anti-realism in truth-value. *Either way, a non-model-theoretic semantics is appropriate* (Shapiro 1997:52) (emphasis added).

But are there any plausible alternatives to the standard model-theoretic semantics that the expressivist can simply pull off the shelf and tweak to accommodate their peculiar semantic and metaphysical constraints? The options appear to be limited, but in the next two sections I will very briefly present two possibilities that the expressivist may wish to consider.

3.2.2 Proof-Theoretic Semantics

One option would be to entertain a *proof-theoretic* semantics of the kind being developed by Schröeder-Heister (2006), Prawitz (2006), and Sundholm (1986), and which finds its roots in Gerhard Gentzen's (1934:68-131). This program is an inferentialist and non-denotational alternative to model-theoretic semantics, and replaces the classical concepts of *truth* (as correspondence) and *truth-condition* with the intuitionistic (and loosely, anti-realist) notions of *proof* and *proof-condition*. In this framework, *inference* displaces *denotation*, and the meaning of each logical connective is determined by its introduction rule in Gentzen's natural deduction system **NJ** (see Gentzen (1934:75) or Troelstra (2000:60)).

Hints of applying similar programs to expressivism are sprinkled throughout the literature, for example in Hale (1993:ft.27), Wedgwood ((1997:81),(2007:56)), and most prominently, in Blackburn (1988:502-503), where he toys with the idea of Prawitz' program in the opening pages of "Attitudes and Contents" while he is canvassing the formal options available for quasi-realism to base its semantics on;

> the view that it is by knowing how to *use* the connectives in proofs that we come to understand them, and hence gain what understanding we have of the truth tables, is perfectly open [...] The whole philosophy of intuitionistic interpretations of the logical constants and of those who give priority to sequent calculi and natural deduction systems opposes the simple assumption that an antecedent understanding of "representative" truth and falsity affords the only road to understanding validity (Blackburn 1988:502-503).

Blackburn of course shirks this route in favor of modeling his quasi-realism in Hintikka's deontic logic, and instead of using *proof* as the key semantic concept in his theory, aims to construct an irrealist moral truth predicate. And it is this choice

that leads him down the path that terminates on one of the two horns of Wright's ((1988:35),(§3.1.3)) dilemma.

So what would be gained by pursuing a proof-theoretic semantics for moral language? For one, adopting it would immediately block a collapse into realism or cognitivism of the sort that plagued Gibbard's normative logic, for here truth and denotation are severed from linguistic items at the formal level, since no relation is established between language and some formal structure representing the world. In addition, rather than attempting to explain the meaning of a sentence in terms of its truth-conditions, which risks violating the Nonfactualism thesis, the meaning of a sentence is determined by how it is *used* in language, in accordance with the natural deduction rules of **NJ**. Proof-theoretic semantics is thus somewhat similar in spirit with the *meaning as use* program in the philosophy of language (see Schröeder-Heister 2006:525). And in this sense, the proposed framework also coheres with the inference-rule account of the conditional developed by Hare and Blackburn (see §3.2.1), where the meaning of the conditional was defined in terms of knowing how to use *modus ponens*, its corresponding elimination rule. In proof-theoretic semantics, a similar relation between the conditional and *modus ponens* is established, albeit with the order of priority reversed;

> The introductions represent, as it were, the 'definitions' of the symbols concerned, and the eliminations are no more, in the final analysis, than the consequences of these definitions [...] By making these ideas more precise it should be possible to display the E-inferences as unique functions of their corresponding I-inferences, on the basis of certain requirements (Gentzen 1934:80-81).

This particular passage from Gentzen is typically taken to be starting point for proof-theoretic semantics, and it leads to the idea that the meaning of a sentence is given *not* by its truth-conditions, but by the ways that it can be *proved* in the system **NJ**. More to the point, the meaning of a sentence is given by its *proof*-conditions, by the set of **NJ**-proofs of it.[14] And since the introduction and elimination rules specify precisely how to use each sentence, plausible accounts of the embedding problem and the validity of the argument **(F-G)** may be forthcoming in this framework.[15]

---

[14]This hints how one might provide a nonfactualist proposition to answer the question posed at the end of Chapter 2 – simply take the set of proofs of a sentence, or its proof-conditions, to be the proposition associated with the sentence.

[15]As an aside, I'll provide a *very* quick and *very* rough sketch of how the Frege-Geach embedding problem might be addressed, stressing that this is a major gloss and is, at best, inconclusive. Let $L =$ "lying is wrong" and $B =$ "getting your brother to lie is wrong." Suppose first that there is a proof π_1 of L. Then the conditional $L \to B$ is obtained by giving a constructive procedure that transforms every proof of L into a proof π_2 of B, in accordance with the proof-conditions of the conditional under the BHK interpretation of intuitionistic logic, and as formalized by the introduction rule $\to I$ for the conditional (see Troelstra (2000:23,36)):

$$\begin{array}{c} [L] \\ \vdots\, \pi_2 \\ \underline{B} \\ L \to B \end{array} \to\text{I}$$

If there is such a constructive procedure, then the embedding goes through. This can then be used to explain the apparent validity of **(F-G)**, which appears, on the surface, to look like $\dfrac{L \to B \qquad L}{B}$ (F-G) . On the current proof-theoretic proposal, **(F-G)** is explained schematically in terms of the justification for the conditional $L \to B$, as on the left below. Then this proof can be *reduced* into the simple form on the right *via* a reduction or normalization procedure (See, for example, Schröeder-Heister (2006:531), Prawitz (2006:516), Martin-Löf (1996:44-48), or Troelstra (2000:27) for variations of this procedure):

$$\begin{array}{cc} \begin{array}{c} [L] \\ \vdots\, \pi_2 \\ \underline{B} \qquad \vdots\, \pi_1 \\ \underline{L \to B \qquad\quad L} \\ B \end{array} \text{(F-G)} & \overset{reduction}{\Longrightarrow} \quad \begin{array}{c} \vdots\, \pi_1 \\ L \\ \vdots\, \pi_2 \\ B \end{array} \end{array}$$

This provides a (very quick and rough) sketch of how the apparent validity of **(F-G)** might be explained proof-theoretically by appealing to the introduction rule of the conditional.

There are some respects then, in which proofs are similar both to the intuitionistic and constructive superwarrant proposed by Lynch (2008:124-125 see also ft.7), and to the commitment tableau proposed by Blackburn (1988) in his logic of attitudes. And while it may seem odd that a formal semantics for natural language is based on the notion of a proof, Schröeder-Heister assures us that

> [...] the term "proof-theoretic semantics" is not any more provocative than Montague's conception of "English as a formal language". Both proof-theoretic semantics and model-theoretic semantics are *indirect* in that they can only be applied via a *formal reading* of aspects of natural language (Schröeder-Heister 2006:527).

Still, the project of proof-theoretic semantics is in its infancy, as compared with the dominant model-theoretic frameworks. And while it is not ready-made to be applied to expressivism, the inferentialist, use-theoretic, and anti-realist themes that are shared by expressivism and proof-theoretic semantics suggest that expressivists ought to keep their eyes on developments within proof-theoretic semantics.

3.2.3 Quasi-Assertions and Justification Conditions

One of the negative results of impaling oneself on the second horn of Wright's (1988:35), (§3.1.3) dilemma was that full-blown assertions fell out of quasi-realism. For if the quasi-realist succeeds in constructing an irrealist truth predicate true$_M$, then to utter a moral sentence like (1) amounts to asserting that the proposition expressed by "lying is wrong" is judged to be true$_M$. And it's hard to see how beliefs (or at least 'beliefs$_M$') do not tag along and displace the non-cognitive desire-like attitudes that are characteristic of the expressivist position.

But perhaps the second horn of Wright's dilemma might be blunted if the expressivist were able to provide an inoffensive account of *assertion* to correspond with the truth$_M$ of a proposition. That is, one might intentionally distance themselves from the inference-rule accounts characteristic of Hale, Blackburn, and the aspiring proof-theoretic semanticist, and instead return to the original target of Geach's test, a more rudimentary speech-act account of expressivism. And one might want to take the lessons learned from our foray into alethic pluralism to admit that moral sentences express truth$_M$-apt propositions, and have truth$_M$-conditions, but that they still differ from ordinary assertions in that sincere utterances of them are meant to perform a *different* kind of speech act than 'genuine' assertion.

In his (1973) discussion of the Frege point, Dummett appears to adopt a view not unlike this. As Wright says, referring to Dummett's (1973:327-48) attempts to embed force indicators into unasserted contexts,

> [t]his suggests, though it is not conclusive, that Dummett was tacitly viewing expressive theories as holding 'expression' to be an illocutionary operation on a thought, just as are assertion, wish, question, and command" (Wright 1988:32).

Whether Wright is correct or not, the theme sketched here still dovetails with Dummett's notion of a *quasi-assertion*, of a particular class of linguistic act, which once recognized, may allow for the second horn of Wright's dilemma to be blunted.

Dummett suggests that assertions of the sort that Geach seeks to refute, which include sincere utterances of moral sentences like (1), may form part of a larger class of *quasi*-assertions that contain as a proper subset the genuine assertions of descriptive sentences. The general idea is that while uttering a moral sentence like (1) does not constitute assertion proper, since it performs the linguistic act of *calling* lying wrong,

it still bears a resemblance to assertion in that its use is governed by certain correctness conditions, and in that expressing assent to it "registers a commitment to some course of action or type of conduct" (Dummett 1973:353). Dummett (1973:357) identifies two 'principal features' that characterize the class of quasi-assertions, including the class of genuine assertions:

qa1 Conventions governing their use are *correct* or *incorrect*. And "[a]n assertion or quasi-assertion requires justification"(1973:357).

qa2 Assertions and quasi-assertions are to be acted upon: "they [...] commit both speaker and hearer, if the latter accepts what is said, to a line of action, linguistic and non-linguistic"(1973:357).

The specifics of the correctness and justification conditions vary depending upon the kind of speech act being performed, though in the case of ordinary assertions Dummett takes them to be given by the speaker *recognizing* that the statement is true, or by recognizing that the proposition expressed by the sentence is true (Dummett 1973:355,359). And for the wider class of quasi-assertions, more general (though still unspecified) *correctness* conditions are said to determine if the speech-act is justified or not.

The appeal of this picture is that it establishes a common ground between the quasi and genuine assertions with the features **qa1** and **qa2**, and yet leaves some room for utterances of moral sentences to have as their primary function a commitment to undertaking an act, of the sort that the moral theorist wants to account for in a formal framework. The crucial differences between moral and descriptive assertions then seem to come from two sources; *first*, the specific details of their respective justification conditions, and *second*, from the particular kinds of actions that they

motivate a speaker and audience to take. And both of these are unique to, and determined by, the specific speech-act that is performed with an utterance of them.

This is a pleasant picture, and I provide it for two reasons. One is that it may provide a hint as to how one might blunt the second horn of Wright's dilemma, though I won't pursue that in detail here. Another reason is that it leads into a specific proposal, due to Dalla Pozza and Garola (1995), which may point to a solution to the Frege-Geach embedding problem. The specifics of the proposal will be omitted, but I will merely remark that it bears some thematic similarities to Dummett's distinction between assertions and *quasi*-assertions, and also with his suggestion that *quasi*-assertions are assigned justification conditions, depending upon what kind of speech act they are used to perform.

3.2.4   Dalla Pozza's Logic of Pragmatics

In Dalla Pozza and Garola (1995) and Dalla Pozza (n.d.), a pragmatic logic for assertions, approvals and disapprovals is developed. Here, Frege's distinction between propositions and judgments (cf. §2.2.1) is exploited to give a two-layered semantic theory. In short, the base layer, or the *propositional layer* of the theory is a classical bivalent logic of *propositions*, with a Tarskian correspondence theory of truth. Here, propositions are assigned truth-values; either *true* or *false.* The general idea is that at the propositional layer, everything is "business as usual."

The innovation here is to postulate a further *proof-theoretic*, or *pragmatic* layer for *judgments* (assertions, approvals, or disapprovals), which operates at a higher level than the propositional layer. Here, logical relations are governed by an *intuitionistic*

logic, and judgments are *not* assigned truth-values, but are assigned *justification values*, in terms of the "intuitive" (Dalla Pozza 1995:82) notion of *proof*: judgments are either *justified* or *unjustified*.

The key feature of the proof-theoretic layer is that it gives a non-truth-functional logic of judgments that allows assertions like $\vdash m$, and disapprovals like $!_w\, l$, to be combined with the (intuitionistic) connectives $\sim_I$, $\cap_I$, $\cup_I$, and $\supset_I$.[16] So, for example, the disapproval $!_w\, l$ and the assertion $\vdash m$ can figure into the antecedent and consequent (respectively) of the conditional

$(6^I) \quad !_w\, l \supset_I \vdash m$

And this conditional is *not* assigned a truth-value, but rather is assigned a *justification value*, which is determined recursively in terms of the justification values of its components.

This might give one pause, for it appears as though (6^I) has violated the Frege point by allowing force-indicators to fall under the scope of a logical connective. But appearances are deceiving, for this conditional is resting safely above at the proof-theoretic layer, which is *not* subject to the Frege point. That is, the Frege point forbids force-indicators to be embedded into *truth-functional* contexts, but such contexts only occur at the base, or propositional layer.

This "logical ascent" not only explains how moral and descriptive judgments can be embedded into complex contexts, but it also provides a quick solution to the problem of mixed moral and descriptive inferences like **(Mix)**. Instead of fumbling about at the propositional layer when attempting to explain why **(Mix)** is valid, we can go

---

[16]Here $m =$ "my mother will be mad at me," $l =$ "lying is wrong," and $!_w$ is the sign of expressive force of *disapproval*, as it was introduced in §1.2.2. Note that the notation that is being used here differs substantially from that used in Dalla Pozza (n.d.).

upstairs to the proof-theoretic layer and look at the corresponding inference $(\mathbf{Mix}^I)$:

$$1. \ !_w \ l$$
$$2. \ !_w \ l \supset_I \ \vdash m$$
$$\therefore \ 3. \ \vdash m$$

Admittedly, the details here are moderately complex.[17] But we can see how if there's a proof that justifies the (moral) judgment $!_w \ l$, and if there's a proof that justifies $!_w \ l \supset_I \vdash m$, then there's a proof that justifies the conclusion $\vdash m$.[18] So $(\mathbf{Mix}^I)$ looks to be valid.

The prospects for this proposal are not yet clear, for while much of the detailed logical work has already been done in (Dalla Pozza 1995), the applications of this logic to the Frege-Geach embedding problem have only been sketched in (Dalla Pozza n.d.). However, it looks to provide at least some partial solutions to the two problems that Blackburn's logic of attitudes foundered on; the problem of mixed sentences, and the problem of negation. And it is also somewhat reassuring to note that the concepts of *proof* and *justification*, which were important in §3.2.2 and §3.3.3 make an appearance here as well.

---

[17]See Dalla Pozza and Garola (1995) and Dalla Pozza (n.d.) for the specific details.

[18]This is clearly a gloss, and even here it skims a bit too much. More properly, corresponding to the second premise something like "if there's a proof that every proof of $!_w l$ can be transformed into a proof of $\vdash m$" should be given, to capture the BHK definition of the conditional.

3.3   Some Concluding Remarks

The initial results yielded by Dalla Pozza's (§3.2.4) pragmatic logic seem rather promising. But if one thing has been established in the course of this thesis, it is that initial appearances tend to be deceiving, and that even the most sophisticated and convincing solutions to the embedding problem tend to buckle under pressure. This was illustrated vividly in §2.1 with Blackburn's logic of attitudes, which yielded to the criticisms of Hale and Unwin (§2.1.3), who showed that there were fatal gaps between the logical formalism deployed by Blackburn, on the one hand, and his preferred quasi-realist interpretation of it, on the other. A similar fate befell Gibbard's normative logic (§2.2), which seemed at first to provide a nice and tight solution to the embedding problem. But Dreier's transformation (§2.2.1) exposed Gibbard's position for what it really was; an indexical theory in disguise.

Even making subtle modifications to the Nonfactualism thesis did not yield any tangible results. Conceding that moral sentences have deflationary truth-conditions allowed for a swift explanation of embedding and the validity of **(F-G)**, but then we were left with a semantic theory drained of all substance, and no plausible account of the meaning of moral sentences (§3.1.1). A commitment to a plurality of truth predicates in §3.1.2 initially seemed like boon, yet when it came time to explain mixed sentences and mixed inferences, the expressivist was at a loss. But perhaps most damaging to the expressivist project was Wright's dilemma (§3.1.3), which still stands unresolved, despite the forays into inference-rule accounts of the conditional in §3.2.1, and suggestions in §3.2.2 that a non-denotational proof-theoretic semantics might lead to a solution. And in the end we're still left with the embedding problem, which has yet to be decisively refuted.

Bibliography

Ayer, A.J. 1952. *Language, Truth and Logic.* Dover.

Beall, J.C. 2000. "On Mixed Inferences and Pluralism About Truth Predicates." *The Philosophical Quarterly.* 50:200. 380-382.

Blackburn, Simon. 1984. *Spreading the Word.* Oxford University Press.

Blackburn, Simon. 1988. "Attitudes and Contents." *Ethics.* 98:3. 501-517.

Blackburn, Simon. 1992. "Gibbard on Normative Logic." *Philosophy and Phenomenological Research.* 52:4. 947-952.

Boghossian, Paul A. 1990. "The Status of Content." *The Philosophical Review.* 99:2. 157-185.

Dalla Pozza, Carlo and Garola, Claudio. 1995. "A Pragmatic Interpretation of Intuitionistic Propositional Logic." *Erkenntnis* 43:1. 81-109.

Dalla Pozza, Carlo. (n.d.) "A *Pragmatic* Logic for the Expressive Conception of Norms and Values and the Frege-Geach Problem." Unpublished manuscript.
`http://siba2.unile.it/ese/issues/ 32/717/DallaPozzaC.pdf`

Dreier, James. 1996. "Expressivist Embeddings and Minimalist Truth." *Philosophical Studies.* 83:1. 29-51.

Dreier, James. 1999. "Transforming Expressivism." *Nous* 33:4, 558-572.

Dummett, Michael. 1973. *Frege: Philosophy of Language.* London:Duckworth.

Edwards, Douglas. 2008. "How to Solve the Problem of Mixed Conjunctions." *Analysis.* 68:2. 143-149.

Fine, Kit. 2001. "The Question of Realism," *Philosopher's Imprint.* 1:1. 1-30.

Frege, Gottlob. 1918-19. "Negation." In McGuinness, Brian (ed.). 1984. *Gottlob Frege: Collected Papers on Mathematics, Logic, and Philosophy.* Basil Blackwell.

Geach, P.T. 1960. "Ascriptivism." *The Philosophical Review.* 69:2. 221-225.

Geach, P.T. 1965. "Assertion." *The Philosophical Review* 74:4. 449-465.

Gentzen, Gerhard. 1934. "Untersuchungen uber das logische Schliessen." In Szabo, M.E. (ed.). 1969. *The Collected Papers of Gerhard Gentzen.* North Holland.

Gibbard, Allan. 1990. *Wise Choices, Apt Feelings: A Theory of Normative Judgment.* Harvard University Press.

Gibbard Allan. 2003. *Thinking How to Live.* Harvard University Press.

Hale, Bob. 1986. "The Compleat Projectivist." *The Philosophical Quarterly.* 36:142. 65-84.

Hale, Bob. 1993. "Can There Be a Logic of Attitudes?" In Haldane, John and Wright, Crispin (eds.). *Reality, Representation and Projection.* Oxford University Press.

Hare, R.M. 1969. *The Language of Morals.* Oxford:Clarendon Press.

Hare, R.M. 1970. "Meaning and Speech Acts." *The Philosophical Review.* 79:1. 3-24.

Hintikka, Jaakko. 1969. "Deontic Logic and Its Philosophical Morals." In Hintikka, Jaakko. *Models for Modalities.* D. Reidel Publishing Company.

Horwich, Paul. 1993. "Gibbard's Theory of Norms." *Philosophy and Public Affairs.* 22:1. 67-78.

Kaplan, David. 1989a. "Demonstratives." In *Themes From Kaplan*, Almog, J. Perry, J. Wettstein, H. (eds.). Oxford University Press

Kaplan, David. 1989b. "Afterthoughts." In *Themes From Kaplan*, Almog, J. Perry, J. Wettstein, H. (eds.). Oxford University Press

Lenman, David. 2003. "Disciplined Syntacticism and Moral Expressivism." *Philosophy and Phenomenological Research.* 66:1. 32-57.

Lynch, M.P. 2008. "Alethic Pluralism, Logical Consequence and the Universality of Reason." *Midwest Studies in Philosophy.* 32:1. 122-140.

Martin-Löf, Per. 1996. "On The Meanings Of The Logical Constants And The Justifications Of The Logical Laws." *Nordic Journal of Philosophical Logic.* 1:1, 11-60.

Prawitz, Dag. 2006. "Meaning Approached via Proofs." *Synthese.* 148. 507-528.

Price, Huw. 1994. "Semantic Minimalism and the Frege Point." In Tsohatzidis, S (ed.). *Foundations of Speech Act Theory: Philosophical and Linguistic Perspectives..* London; New York: Routledge.

Searle, John. 1969. *Speech Acts: An Essay in the Philosophy of Language.* Cambridge University Press.

Schroeder, Mark. 2008. "Expression for Expressivists." *Philosophy and Phenomenological Research.* 76:1, 86-116.

Schröeder-Heister, Peter. 2006. "Validity Concepts in Proof-Theoretic Semantics." *Synthese.* 148. 525-571.

Schueler, G. F. 1988. "Moral Modus Ponens and Moral Realism." *Ethics.* 98:3. 492-500.

Shapiro, Stewart. 1993. "Modality and Ontology." *Mind: New Series.* 102:407. 455-481.

Shapiro, Stewart. 1997. *Philosophy of Mathematics: Structure and Ontology.* Oxford University Press.

Sinnott-Armstrong, Walter. 1993. "Some Problems for Gibbard's Norm-Expressivism." *Philosophical Studies.* 69. 297-313.

Sinnott-Armstrong, Walter. 2000. "Expressivism and Embedding", *Philosophy and Phenomenological Research.* 61:3. 677-693.

Stoljar, Daniel. 1993. "Emotivism and Truth-Conditions." *Philosophical Studies.* 70. 81-101.

Sundholm, Göran. 1986. "Proof Theory and Meaning." In Gabbay D. and Guenthner F. (eds.). *Handbook of Philosophical Logic: Vol III* (pp. 471-506). Dordrecht; London: Springer.

Sundholm, Goran. 2002. "A Century of Inference." In Gärdenfors, P. (ed.). *The Scope of Logic, Methodology and Philosophy of Science: Volume II* (pp.565-580). Dordrecht: Kluwer.

Tappolet, Christine. 1997. "Mixed Inferences: A Problem for Pluralism About Truth Predicates." *Analysis.* 57:3. 209-210.

Tappolet, Christine. 2000. "Truth Pluralism and Many-Valued Logics: A Reply to Beall." *The Philosophical Quarterly.* 50:200. 382-385.

Tennant, Neil. 1997. *The Taming of the True.* Oxford: Clarendon Press.

Troelstra, A.S. and Schwichtenberg, H. 2000. *Basic Proof Theory* (2nd ed). Cambridge Tracts in Theoretical Computer Science.

Unwin, Nicholas. 1999. "*Quasi*-Realism, Negation and the Frege-Geach Problem." *The Philosophical Quarterly.* 49:196. 337-352.

Unwin, Nicholas. 2001. "Norms and Negation: A Problem for Gibbard's Logic." *The Philosophical Quarterly.* 51:202. 60-75.

van Roojen, Mark. 1996. "Expressivism and Irrationality." *The Philosophical Review.* 105:3. 311-335.

Wedgwood, Ralph. 1997. "Non-Cognitivism, Truth and Logic." *Philosophical Studies.* 86. 73-91.

Wedgwood, Ralph. 2007. *The Nature of Normativity.* Oxford: Clarendon Press.

Wright, Crispin. 1988. "Realism, Antirealism, Irrealism, *Quasi*-Realism." *Midwest Studies in Philosophy.* 12. 25-49.

Wright, Crispin. 1992. *Truth and Objectivity.* Cambridge Mass: Harvard University Press.

Wright, Crispin. 1995. "Truth in Ethics." *Ratio (New Series)* 8:3. 209-226.