How to Do Things with Norms: A Speech Act Approach to Metanormative Theory

Andrew Dale Morgan
Charlottesville, Virginia

Bachelor of Arts, Lee University, 2009
Master of Arts, Northern Illinois University, 2011

A Dissertation presented to the Graduate Faculty of the
University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Department of Philosophy

University of Virginia
May, 2017

Contents

Abstract

When we use normative language—terms like ‘good,’ ‘vicious,’ or ‘beautiful,’—what exactly are we up to? On the one hand, it feels as though we are trying to express our beliefs to one another. On the other hand, it also feels as though we are trying to express other attitudes like praise, condemnation, or awe. This creates a puzzle for philosophers: how do we adequately capture both aspects of our normative language? In this dissertation I argue, first, that existing strategies for resolving this puzzle fail, either because they tie the expression in question too closely to the semantics of normative terms or because they tie it too loosely to the features of a speaker’s context. I then present my own positive view, the hybrid speech act theory. The central insight of my view is that when we make normative claims, we are making use of distinct (and hitherto unrecognized) types of speech acts. What sets normative speech acts apart is that their constitutive sincerity conditions require speakers to possess both cognitive and motivationally efficacious states. My dissertation concludes with a discussion of what we are up to when we have normative thoughts. By taking a closer look at the phenomenon of inner speech, I demonstrate how my normative speech act theory may be used to explain normative thoughts as well.

UNIVERSITY of VIRGINIA
GRADUATE SCHOOL of ARTS & SCIENCES

Graduate School of Arts and Sciences Enrolled Student Office
P.O. Box 400772
Charlottesville, VA 22904-4773
http://graduate.as.virginia.edu/
gsasregistrar@virginia.edu

FINAL EXAMINATION FORM

This form is to be completed by a student's committee and department to indicate that the student has passed the final defense of their thesis/dissertation, written or oral exams. This form must be scanned and emailed to the Enrolled Student Office by the appropriate deadline: **November 30** for December, **April 30** for May, and **July 30** for August graduation.

Last Name: Morgan | First Name: Andrew | University/SIS ID: 2 2 6 1 1 2 4

Program: Philosophy | Date Final Examination Passed: 3/24/17

Master's Final Examination: A candidate must achieve satisfactory standing in a final examination (oral, written or both) conducted by two or more graduate faculty members designated by the candidate's department.

PhD Final Examination: This committee, chaired by the primary advisor, will consist of a minimum of four tenured or tenure-track members of the faculty of the Graduate School of Arts and Sciences. One member of the committee will serve as a representative of the Graduate School of Arts and Sciences to affirm that the student has been assessed fairly and with due rigor. This representative is appointed by the student's director of graduate studies and must hold a primary appointment outside of the student's department. This representative may be drawn from the tenured or tenure-track faculty of other graduate schools at the University, but must hold a Ph.D. A director of graduate studies may petition to permit a reader from outside the University who holds a Ph.D. to serve as one of the four core members of the committee by providing the associate dean with the reader's CV and a statement regarding the reader's particular suitability for the committee. This external reader may not serve as the representative of the Graduate School. Once these minimum requirements have been met, additional committee members from within the University or other institutions may be added.

Committee	Name	Signature	Department
Primary Advisor	Talbot Brewer		Philosophy
Committee Member	Harold Langsam		Philosophy
Committee Member	Rebecca Stangl		Philosophy
GSAS Representative	Amrisha Vaish		Psychology
Committee Member (optional)	DANIEL BOISVERT		UNC CHARLOTTE PHILOSOPHY

Approved Title of Doctoral Dissertation or Master's Thesis if Required

How to Do Things with Norms: A Speech Act Approach to Metanormative Theory

ANTONIA LoCASCIO
Graduate Advisor (if no thesis, Graduate Studies Director) (Print)

Signature

3/24/17
Date

Talbot Brewer
Department Chair or Representative (Print)

Signature

3/24/17
Date

***If Advisor is the Graduate Director, Chair Signature is required. If Chair is advisor, Graduate Director Signature is required.*

This examination is in fulfillment of the requirements as described in the Record for the appropriate degree.

Please scan and email this form to gsasregistrar@virginia.edu

Acknowledgments

This dissertation is the record of only one half of a conversation. Over the past six years there have been many people on the other side of that exchange without whom it is hard to imagine my thoughts taking coherent shape at all, much less expression in the words that follow. That conversation began in courses at the University of Virginia taught by Rebecca Stangl and Mitch Green, who encouraged some of my earliest forays into the worlds of metaethics and pragmatics respectively. I am also indebted to Mitch for the many subsequent conversations that prodded me to expand both the scope of the project (to varieties of normative discourse beyond the ethical) and its depth (beyond social speech to inner speech). Thanks to my fellow graduate students who have added their voices to the discussion and many of whom read several drafts of various portions of the text, especially Matthew Adams, Ralph DiFranco, Matt Duncan, Corin Fox, Bill Hasselberger, Derek Lam, Nick Rimell, and Stacie Thyrion. Thanks to Loren Lomasky for his advice and friendship, and for pushing me to never be satisfied with anything subpar—either in philosophy or in food.

Many thanks to Dorit Bar-On, Michael Cholbi, David Copp, and Brie Gertler for their generosity of time in careful reading, casual chats, and correspondence on crucial aspects of my arguments. Added thanks to David Copp for introducing me to Dan Boisvert, and thanks to Dan for his tireless support and comments on the entire document. There is a unique kind of encouragement that can only be provided by someone with a stake in a rival answer to the same philosophical question, and I am indebted to Dan for that irreplaceable source of reassurance. Thanks to Tal Brewer, my advisor. More so than anyone else your voice can be detected in these pages, silently prodding me to examine the issues at hand in a way that is more careful, serious, and sincere. From our countless conversations I have come to appreciate in a new way the power of philosophy as an activity both for self-understanding and cooperative discovery.

Finally, love and gratitude to my family. To my parents Dennis and Cynde for their patience and faith in my ability to succeed and to my sister Heather for (mostly!) putting up with my earliest experiments in cool-headed rational argument. All my love to my wife, Anne. You have walked beside me through the most difficult parts of this process. Without you I would not be the person capable of these thoughts. I also wouldn't have been surrounded by so many wonderful, comforting cats!

CHAPTER 1

WHY SHOULD WE WANT A HYBRID METANORMATIVE THEORY?

Part I: Introduction

Section 1: What is the Scope of Normativity?

What are we doing when we think and talk about norms? The form of the question itself suggests the beginning of an answer: we are *doing* something. That is to say, when we think and talk about norms we are first and foremost engaged in activities. In the case of normative talk, it should be obvious enough that we are engaged in conversations. In the case of normative thought, it is a bit less clear. We often seem to have beliefs about norms. Many times we also have internalized dispositions to follow certain norms. Most of the time these beliefs and dispositions exist below the surface of awareness, and it can take some work to make them conscious—even if (especially if!) they have been influencing the way we view the world and the way we act for a long time. In contrast with these oft-concealed dispositions, when I ask what we are doing when we think about norms I am asking about an activity we engage in at the level of conscious thought.

Thinking and talking about norms can often have a causal influence on the shape our future dispositions and actions take. If it didn't, what would be the point of these activities? What would be the point of having norms in the first place? After all, what are normative concepts for if they aren't tools we can use to uncover, understand, give voice to, and ultimately influence dispositions and actions? Perhaps this is a good way to delineate just what we mean to pick out when we speak about norms. If any of the so-called "norms" of epistemology, rationality, aesthetics, prudence, morality, etiquette, etc. don't in some way find their purpose tied up in the goal of understanding and shaping the way we act then perhaps it is time to stop calling them norms. This is how I'll restrict the scope of my project. The normative realm is the realm that we draw upon when we attempt to answer Socrates' question "How should one live?" where "*should*" as Bernard Williams says "is simply *should*."[1]

To see if this makes sense, let's take a look at one of the harder cases. Suppose we find ourselves discussing Socrates' question together. You bring up as a consideration the fact that a great deal of the music of J.S. Bach is quite beautiful. Now suppose my reaction is that your claim is entirely irrelevant to our discussion. Is the concept of BEAUTY functioning as a *normative* one for me?

I don't think it is. And I think that this is an indication that I am losing my grip on the concept of BEAUTY altogether—what it means to call something beautiful, or to think it beautiful. To judge that something is "beautiful" presupposes, among other things, that the fact that it is beautiful is relevant to a complete understanding of how one should live,

[1] Williams (1985) p. 5. That is to say, "no prior advantage is built into the question for one kind of reason over another" (Ibid. p. 19). For a sympathetic approach to picking out the space of normativity see Gibbard (1990), p. 33.

all things considered. One should, all else being equal, live in such a way as to appreciate beauty when it is encountered. To disagree with this is to say that BEAUTY is not a normative concept—which is, I submit, to say that there is no such thing as beauty at all.

I might have another reaction that at first seems quite similar, but the distinguishing of which will be instructive. That is, I might say that your bringing up of the beauty of Bach's music is *inappropriate*. For instance, I might find myself in a situation that demands confronting and appreciating the ugliness of certain aspects of life. In such a case, it could very well be that dwelling on the beauty of certain pieces of art is quite an inappropriate thing to do while grappling with the question of how I should live at that precise moment.

But that is not to say that Bach's music isn't beautiful. It is rather to recognize that BEAUTY is a thick concept, a concept that tells you something more about the sort of thing something is beyond just that it has aesthetic value or significance. Just as the virtuous person can find themselves in a situation where justice is not the trait that is called for, any one of us may find ourselves in a situation when reflection on beauty is not what is called for. Sincerely acknowledging something as beautiful demands accounting for in the future contours of our lives, it is a more than merely cognitive experience. In fact, it is *partly because* wholehearted recognition of beauty involves affective states—and sometimes even motivating dispositions—that dwelling on the beautiful could sometimes be the wrong thing to do. The presence of this possibility casts no doubt at all on BEAUTY's *normativity*.

What we are doing when we think and talk about norms is we are engaging in an activity that is aimed at answering the question "how should one live?" When we talk about norms we are engaging in this activity *interpersonally*. When we think about norms we are

engaged in this activity *intrapersonally*. If the activity we are engaged in isn't aimed at answering this question, then whatever else it may be it in some sense *falls short* of fully normative engagement with the problem at hand. Acknowledging various answers to this question *wholeheartedly* in some way implicates commitments on both cognitive and motivationally efficacious levels—at least at first blush.

Section 2: What is a Hybrid Metanormative Theory?

A metanormative theory is just a theory that attempts to explain what it is we are doing when we think or talk about norms. Metanormative theories that attempt to incorporate *both* cognitive and motivational/affective elements into these explanations have come to be known as "hybrid." In one sense such theories have been around for some time. Moral psychologists from Plato to Hume have been grappling with the proper way to understand the way our cognitive and affective faculties interact in ethical judgment making. More recently, an explicitly metanormative discussion has been going on since Michael Smith's *Moral Problem* gave new life to the puzzle in the 1990s. Over the past 30 years, philosophers have attempted to solve the problem by choosing a side (cognitive *or* moving) and attempting to explain away the appeal of the other side. The implicit assumption has been that no theory can have its cake and eat it too. The recent emergence of the hybrid approach challenges the status quo in two primary ways: first by working towards a unified meta-theory of all normative domains (not just the moral), and second by explicitly seeking to put the cognitive and motivational/affective dimensions on an even playing field for the first time.

Hybrid theories do this by attempting to analyze speakers who make public normative claims as thereby expressing both cognitive and motivational (or affective) states, *and/or* by analyzing normative mental judgments as involving both cognitive and motivational (or affective) elements.[2] In other words, if a hybrid metanormative theory is attempting to get at what we are doing when we **talk** about norms, it will claim that we are (among other things) expressing both what we believe and how we are disposed to act or feel. If a hybrid metanormative theory is attempting to get at what we are doing when we **think** about norms, it will claim that we are (among other things) attending to our conscious beliefs and motivations or feelings. For the rest of this chapter I will refer to a subject's being "moved" or her "moving attitudes" to capture the range of psychological dispositions that involve motivations and feelings (desires, attitudes, emotions, effective states, etc.).

More precisely, a hybrid metanormative theory will try to give an account of what we are doing when we talk and think about norms in a way that is consistent with either or both of the following two theses:

1. A **speaker** who wholeheartedly speaks aloud a normative statement has in some way communicated to her audience both that she consciously believes what she says and that she feels moved in accordance with what she says.

2. A **thinker** who wholeheartedly thinks a normative thought consciously believes what she has thought and feels moved in accordance with what she has thought.

It should be clear that these theses as they stand are neutral with respect to what exactly it is for speakers to "make a normative claim" and for thinkers to "make a normative judgment." These theses are also neutral with respect to exactly which concepts and terms

[2] These are (more or less) the criteria used by Fletcher and Ridge in their (2014).

are the normative ones, and about just what it may look like to "feel moved" in accordance with them—not to mention how this "accordance" may differ significantly depending on which normative term or concept is at stake.

For example, if I **think** to myself sincerely and wholeheartedly that the music of J.S. Bach is beautiful then according to a hybrid metanormative theory I must *both* believe that his music is beautiful and have had some sort of corresponding affective reaction to his music. In contrast, if I **say** to you in all seriousness and without qualification that environmental apathy is a vice that is destroying our planet then—according a hybrid metanormative theory—I have communicated to you *both* that I believe that environmental apathy is a vice and that I am at least somewhat motivated to work against the existence of such a vice.

The hybrid metanormative project is prefaced on the assumption that consistency with a broadly Humean theory of mind is a theoretical virtue. This is not to say that all hybrid views *imply* Psychological Humeanism, but only that the falsity of Humeanism would remove one of the motivations for seeking such a view. By "Psychological Humeanism" I mean here just the minimal view that (1) cognitive and moving states are not identical, and (2) no cognitive state is *necessarily* conjoined with a moving state. If Humeanism were false, then accounting for the descriptive and practical elements of normativity would be a simple matter of appealing to a more complex theory of mind. Unfortunately, arguing for the truth of Humeanism is beyond the scope of this project.[3]

[3] That is, such a discussion is beyond the scope *except* insofar as demonstrating the clear and coherent possibility of a satisfying hybrid view might remove possible objections to the Humean view.

However, there is another benefit to remaining consistent with Humeanism, which stems from the fact that it essentially consists in two *negative* theses. A hybrid metanormative theory that is able to *avoid assuming* that cognitive and moving states are sometimes identical, or that they are ever necessarily conjoined, has a much broader appeal than a theory that does involve such commitments. Such a theory is consistent with a wide range of both Humean and anti-Humean theories of mind. This is an advantage because most people, when they think and talk about norms, do not themselves presuppose a sophisticated theory of mind. If the hybrid metanormative theorist wishes to capture what most people are doing, then it is better for the explanation offered not to depend entirely on one particular (and perhaps controversial) theory of mind.

The rest of this chapter proceeds straightforwardly in two main sections. Part II gives reasons for wanting the cognitive element in our metanormative theory. It argues that the most natural way to account for our talk and thought about norms takes sincere wholeheartedness to indicate belief in what we say and judge. Part III gives reasons for wanting the moving element in our metanormative theory. It argues that the most natural way to account for our talk and thought about norms takes sincere wholeheartedness to indicate at least some level of motivation or feeling appropriate to what we say and judge. The conclusion of the chapter is that, given that we want both the cognitive and motivational/affective elements in our metanormative theory, what we really want is a hybrid metanormative theory. It closes with a brief outline of the remaining chapters of the dissertation.

Part II: Metanormative Cognitivism

Section 1: The Pre-Theoretical Advantage of Cognitivism

Suppose your neighbor says to you in all seriousness, "We really should get together more often." How might she respond if you were to reply "Sure, but do you *believe* that we should?" She wouldn't merely answer your question with a resounding "yes!" Wouldn't she be puzzled by the fact that you asked the question in the first place? That is, wouldn't she take herself to have already given you the information you are asking for?

In fact, it would be so obvious to her that she had already expressed to you the information that you purport to be seeking, that your question could easily cause offense: "What do you mean 'do I *believe* that'? I just said that we should!" Your question would, more likely than not, be reinterpreted as a challenge to the sincerity of her original statement, or perhaps even an expression of misgivings about the truth of the original statement. In short, everyday conversational pragmatics seem to rely on the presumption that speakers who make normative claims (at least when they do so in a serious way) express that they believe what they are saying—just as speakers who make non-normative assertions do.

Notice that this expression of belief comes across despite the fact that it isn't entirely clear what sort of normative force your neighbor is conveying when she says the two of you really "should" get together more often. Is it a moral "should" or a prudential one? Or is she making a claim about etiquette? Your neighbor herself may not know, and a careful analysis might ultimately reveal that an overlapping combination of different

dimensions of normativity is involved. This indicates that *whether belief is expressed doesn't depend on the sort of norm being invoked*.

Notice also that your neighbor appears to be doing more than merely evincing her preference for getting together more often, making a recommendation, or issuing a command—though these sentiments may be expressed as well (let's leave that issue for the next section). She at least *seems* to be putting a consideration forward as relevant to answering Socrates' question. You can't count as fully disagreeing with her simply by taking issue with her internal states, or by acting in a way that is contrary to her statement. The only way, really, to disagree is by saying that what she says isn't true—that her claim doesn't add anything to the mutual understanding of how the two of you should live.

Perhaps it isn't good to make *too much* of everyday conversational pragmatics. And perhaps we shouldn't *always* take speakers at their word when they say they believe things. Nevertheless, I think it is important to point out that there is a cost to denying metanormative cognitivism. Not just a theoretical cost, but a social cost as well: to give up on cognitivism is to refuse to cooperate in a fairly basic way with other people.

There may be ways around this, but it is significant that such theoretical gymnastics count as *ways around*. That is to say, a denier of cognitivism may be able to make sense of our ordinary ways of speaking, but what they are doing when they do so is discharging a burden—they are overcoming an obstacle, solving a problem. Non-cognitivism, taken simply as the rejection of cognitivism, is not the natural way we read our normative interactions with others. It is a position that may be able to explain away our intuitions as confused, but even if this explanation is given in a very elegant and satisfying way it is

only worth going in for in the first place if there is some other independent theoretical benefit to doing so.

In other words, the best case that could be made *for* cognitivism would be to show that there is no convincing case *against it.* The burden of proof is on the non-cognitivists to say why we should even attempt to reinterpret our natural ways of speaking, thinking, and interacting. In the next section I'll clarify what such reasons would have to look like, and why in the context of our overall project there really aren't any.

Section 2: The Post-Theoretical Advantage of Cognitivism

To begin, let's be clear about exactly what the theses are to be defended. First of all, I will understand metanormative cognitivists to be committed to the following:

> **Psychological Cognitivism (PC):** Whenever a speaker wholeheartedly utters a normative claim, she expresses that she believes what she says AND whenever a thinker wholeheartedly makes a normative judgment, she believes what she has judged.

Notice that nowhere in Psychological Cognitivism is it claimed that expressing beliefs is *all* that normative speakers are doing, nor is it claimed that believing is *all* that normative thinkers are doing. Also, in this context 'judgment' is meant to be a neutral term, one that picks out *whatever it is* that thinkers are doing—the mental act analogous to seriously uttering a normative claim aloud. Notice also that on its own PC has nothing to say about whether the propositions that are the objects of these beliefs are truth assessable. However, I take it that for the vast majority of theorists in the vast majority of cases the mere fact *that a proposition is the object of a belief* straightforwardly implies that it is either true or false. To those theorists who think that there are cases in which we are able to form beliefs

that are not truth apt I'm really not sure what to say. Perhaps I will just admit that, were I convinced that our normative talk or thought had succumbed to such a strange fate, I would be more confident of the strength of its connection to belief than I would be of its connection to truth assessability.

Setting such possibilities scenarios aside, it is best to go ahead and commit the cognitivist to a semantic thesis as well:

> **Semantic Factualism (SF):** Properly formed declarative sentences in which normative terms are used express propositions that are truth assessable.

Semantic Factualism is a natural view for a proponent of PC to hold. It is also fairly natural to think that the content of these propositions may be specified *disquotationally*. For example, the sentence 'Mountains are beautiful' expresses the proposition that mountains are beautiful, and this is what explains the fact that it is a good English translation of the German sentence 'Berge sind schön.'[4]

In fact, committing oneself to this understanding of SF is what helps the cognitivist to navigate Frege's abyss in such a straightforward and intuitive way. That is to say, among other things, that the proponent of SF is able to count the following argument as being deductively valid:

1. Mountains are beautiful.
2. If mountains are beautiful then the Grossglockner is beautiful.
3. The Grossglockner is beautiful.

[4] I take it that, at least so far as we have gone, SF is neutral with respect to most theories of meaning, reference, truth, and the nature of propositions. For a well put discussion of why it is a virtue for a theory of normative thought and talk to remain semantically conservative see Bar-On et al. (2014), pp. 226-231.

Of course, many who deny SF have ways of doing this, or at least of explaining away why we *think* such arguments are deductively valid. But again, such are ways of making the best of a less than ideal situation.

From now on I will use the term 'cognitivism' to pick out the conjunction of Psychological Cognitivism and Semantic Factualism. Why might someone possibly deny cognitivism? Historically, there are two primary reasons. The first consists in strong independent arguments against moral realism, together with some additional assumptions. The second involves the claim that cognitivism is unable to fully account for the practicality of normative thought and talk. These two sorts of reasons often come together, so it will be helpful to pull them apart by looking at the case of one of the first philosophical non-cognitivists: A. J. Ayer.

Ayer's statement of emotivism about ethics was one of the first explicit denials of cognitivism. And there are ready explanations for this rejection: a unique combination of logical empiricism and Moorean nonreductionism. Before leaving to work with the Vienna Circle, Ayer read G.E. Moore's *Principia* and was utterly convinced by its Open Question Argument.[5] That is to say, Ayer believed that normative terms could *not* be analyzed as referring to natural (or as Ayer would say, "observable") properties. Add this to the positivist criterion of meaning—a sentence is truth-evaluable only if it could be confirmed or disconfirmed by empirical observation—and you've got a simple recipe for the denial of cognitivism. Notice that neither Ayer's positivism nor his Moorean intuition would be, on its own, sufficient for denying cognitivism.

---

[5] See Rogers (1999), pp. 53 and 121.

Though logical empiricism has long since fallen out of fashion, we can still see in Ayer's descendants a combination of similar commitments that lead them to deny cognitivism. That is, they are persuaded by Moore so far as the Open Question Argument goes, but jump off the wagon before arriving at Moore's intuitionist moral realism. And though empiricism has had its share of difficulties, various forms of naturalism and physicalism are alive and well. Thus, if normative terms *don't* refer to natural properties one might find oneself embracing a dilemma: either normative terms are in the business of referring or they aren't. The former path leads the empirically-inclined metaethicist to either error theory or constructivism, while the latter leads to non-cognitivism.

So the first major point to appreciate is that one can be a committed anti-realist and still be a cognitivist. Unless you are committed to *both* normative non-reduction a la the Open Question Argument *and* an unreasonable criterion of truth-evaluability, the only way for you to object to cognitivism is by saying that no cognitivist theory (realist, fictionalist, constructivist, or otherwise) can adequately account for the practical role that normative thought and talk play in our lives. It should by now be clear that the anti-realist objection to cognitivism really just collapses to the practicality objection. This leads us to the second major point that needs to be appreciated, which is that the practicality objection is really just the objection that there *couldn't be a satisfying hybrid metanormative view.*

Let's take stock. At the end of the day there are two real reasons why someone would reject cognitivism: (1) commitment to the irreducibility of normative terms together with an implausibly restricted criterion of truth-assessability, or (2) skepticism that any cognitivist view can really do justice to the practicality of normative thought and talk. I have nothing, really, to say to proponents of (1), the last holdouts of the logical positivist

program. To those suffering from (2), I would say that you still have every reason in the world to *want* a hybrid metanormative theory. Whether a coherent and plausible such view may be constructed will be left to other chapters to show.

Part III: Metanormative Practicality

Section 1: A Few Clarifications

In this section I will argue that another thing we really want out of a metanormative theory is a commitment to the following thesis:

> **Practicality (P):** Whenever a speaker wholeheartedly utters a normative claim, she communicates that she has some moving attitude that is in accordance with what she says AND whenever a thinker wholeheartedly makes a normative judgment, she has some moving attitude that is in accordance with what she has judged.

Just like Cognitivism, Practicality has nothing to say about what other sorts of things normative speakers and thinkers may be up to. This is significant: there is no reason on the face of things to expect PC, SF, and P to come into conflict with one another.

However, remember that one of the primary motivations for pursuing the hybrid metanormative approach in the first place was a desire to remain consistent with Psychological Humeanism. Historically, it has been precisely this cocktail of views—cognitivism, practicality, and Humeanism—that has been thought to create problems. Take for example Michael Smith's famous paradox:

A. Moral judgments of the form 'It is right that I Φ' express a subject's beliefs about an objective matter of fact, a fact about what it is right for her to do.
B. If someone judges that it is right that she Φs then, *ceteris paribus*, she is motivated to Φ.
C. An agent is motivated to act in a certain way just in case she has an appropriate desire and a means-ends belief, where belief and desire are, in Hume's terms, distinct existences.[6]

It may not at first be obvious why someone would be tempted to think that these three theses are inconsistent. Smith brings out the tension by illicitly adding to cognitivism (statement A) the assumption that "*the* state expressed by moral judgment is belief" which he then goes on to presume really just amounts to the view that moral judgments *are* beliefs. Nowhere in my definition of Cognitivism (nor in Smith's original statement A) is such an identification made. So if one wants a hybrid metanormative theory to be neutral with respect to Psychological Humeanism, all one need do is abandon Smith's chain of assumptions. This is an important lesson to keep in mind: in order for hybrid metanormative theories to have a fighting chance, they must affirm a version of Cognitivism that does not identify normative judgments with beliefs. Importantly, this *does not mean* that the hybrid theorist must deny the possibility of *having* a belief with normative content. This lesson will return to play a significant role in Chapter 4.

Another observation that is crucial to bring up at the outset of a discussion of the practical force of normative thought and talk is that different sorts of norms can often at least appear to conflict with one another. For example, it is not patently absurd to think that sometimes it is prudent to form beliefs that are epistemically unjustified, morally right to

[6] Smith (1994), p. 12. A is meant to stand for something like cognitivism, while B is a version of internalism (a close cousin of Practicality) and C is one way of stating Psychological Humeanism.

say something that is impolite, or beautiful to create something that purports to represent as possible a logically *impossible* state of affairs (irrational).

To see this, simply imagine an athlete preparing for a competition, a musician preparing for a concert, or a candidate preparing for an interview. Suppose that their lives will go better for them, overall, the more successful they are in these endeavors. Suppose further that they would be entirely unjustified in believing that their performance will be flawless. It isn't at all hard to conceive of cases in which, all else being equal, the individual who believes they will perform flawlessly will be *more* successful than the individual who only forms the epistemically justified belief ("I most certainly *will not* perform flawlessly").

Other examples are easy to come by: cases in which the moral thing to do is to abruptly exit a friend's house because they are making offensive remarks, rather than politely laugh along; or paintings that beautifully portray impossible objects by means of optical illusions; or times when you must choose between being benevolent and being just; or situations that pit the goodness of knowing the truth against the goodness of a sound night's sleep. Now at this point, for many, red flags may begin to go up. Of course, one could argue for the unity of the virtues. Or you might think that ultimately all the different sorts of value reduce to a single normative domain. Some are even committed to the idea that considerations from one normative domain may be so overriding as to silence demands made from other domains.

I do not deny any of this. My aim here is simply to point out that when we are preparing to examine our intuitions about the practicality of normative thought and talk, it is necessary to keep in mind that sometimes our own responses to the normative features

of a situation may be at loggerheads with one another. For purposes of clarity we will need to speak somewhat unrealistically of "all else being equal," though all else is hardly ever equal. In the next sub-section we will survey some reasons for thinking that Practicality is true. As we do so, it will be important to remember that very often the fact that we don't feel the pull of one sort of reason in a particular case casts no doubt at all on the Practicality thesis, but rather indicates that we are being carried away by countervailing reasons that have been smuggled in through the back door.

Section 2: In Favor of Practicality

> *Socrates*: And, in your opinion, do those who think that [evil actions] will do them good know that they are evils?
> *Meno*: Certainly not.
> *Socrates*: Is it not obvious that those who are ignorant of their nature do not desire them; but they desire what they suppose to be goods although they are really evils; and if they are mistaken and suppose the evils to be good they really desire goods?
> *Meno*: Yes, in that case.[7]

> If morality had naturally no influence on human passions and actions, 'twere in vain to take such pains to inculcate it; and nothing wou'd be more fruitless than that multitude of rules and precepts, with which all moralists abound. Philosophy is commonly divided into speculative and practical; and as morality is always comprehended under the latter division, 'tis supposed to influence our passions and actions, and to go beyond the calm and indolent judgments of the understanding. And this is confirm'd by common experience, which informs us, that men are often govern'd by their duties, and are deter'd from some actions by the opinion of injustice, and impell'd to others by that of obligation.[8]

> Moral understanding requires that those who would claim to have it should be serious respondents to morality's demands. Someone who cannot be responsive to morality's demands is one for whom morality has no reality. The 'reality' of moral value is inseparable from the reality of it as a claim on us, and serious responsiveness to that claim is internal to the recognition

[7] Plato, *Meno* 77d–e (Jowett 1952, p. 178).
[8] Hume, *Treatise* 3.1.1.5 (Norton and Norton 2000, p. 294).

> of its reality. (That is the element of truth in 'non-cognitivism', particularly emotivism.) To understand, for example, what we are contemplating if we are contemplating murder, we must understand what it is to be a murderer, which is to understand rightly in what way it matters. I discuss this more fully when I turn to more detailed exploration of Socrates' claim that it is better to suffer evil than to do it and of its relation to his claim that we cannot do evil knowingly. For the present, it is sufficient to say that the inner disintegration characteristic of madness makes impossible the serious responsiveness to the claims of morality that is internal to the recognition of their reality. That is one reason for not calling remorse a mere feeling, or a mere attitude. Remorse is, amongst other things, a disciplined remembrance of the moral significance of what we did.[9]

> Anna says that capital punishment is morally wrong, yet whenever she hears that a criminal is about to be executed, she seems indifferent. There is never any indication that she disapproves of capital punishment, other than the fact that she says it is wrong. When we challenge her, she continues to insist that capital punishment is wrong, but she agrees that she has no negative attitude toward it. She denies that she disapproves of it. Something appears to be amiss. Her assertion seems to be at odds with the indifference that she displays and avows.[10]

We said at the outset that one clear way to delineate the scope of thinking and talking that we would be willing to count as normative is that it is in some way aimed at answering what Bernard Williams calls "Socrates' question": how should one live? What I believe each of these four authors is trying to get at, in their own way, is that you can't really appreciate some notion as *counting as a potential answer* to this question if you don't, at the same time, recognize that *in offering it* the answerer "sticks his or her neck out"—so to speak.

---

[9] Gaita (2004), p. 59.
[10] Copp (2009), p. 168.

One way to bring out this idea is by attending to the phenomenon of normative Moorean absurdity.[11] Here are just a few examples, artificial snapshots of hypothetical conversations in which something strikes us as having gone amiss:

> **Case 1:** I do think that, given her situation, the right thing for her to do is return the money. I mean, that *is* what a truly virtuous person would do. All the same, I don't think I would approve at all of her fulfilling her obligations.
>
> **Case 2:** Did you know that he is a climate change skeptic? He is an intelligent guy and he's seen all the evidence—still he persists in believing that the activities of humans have had absolutely no impact on the rise in global temperatures since the industrial revolution. But of course, I don't really disapprove of his willful ignorance.
>
> **Case 3:** Don't you think it would be in his best interest to wait on buying the house for now? The market is sure to go down a lot more in the coming months, and rental prices are still so affordable. In the long run, he'll surely regret his decision—there's really no good reason at all to rush into it. Nevertheless, if it were my decision to make I wouldn't hesitate to do the exact same thing.
>
> **Case 4:** Isn't her performance one of the most fitting and sensitive interpretations of the piece you've heard? I don't want to lay it on too thick, but I really couldn't imagine it coming off more poignantly, or more beautifully. She really does have an instinct for Chopin—an ability to express gravitas without becoming trite or melodramatic. Even so, listening to her play really produced no effect on me and I couldn't care less whether I heard her again.

The overall pattern should be clear by now. When speakers commit themselves to particular evaluations within some normative domain or other, but then contrast these commitments against a complete lack of any motivation or feeling in accordance with them, the audience is left puzzled.

---

[11] Inspired by Michael Cholbi's "moral Moore-paradoxical propositions." See Cholbi (2009).

This is not to deny that there are ways of filling out these cases that might ease our puzzlement. We might learn more about the speaker, such that the disavowal of moved feeling is in some sense comprehensible to us—even held alongside their strong normative evaluation. In case 4 we might learn that the speaker has a strong personal dislike for the performer she is praising. In case 2 we might learn that the speaker cares greatly for the person she is criticizing, and doesn't hold her friend's epistemic failings against him. In case 1 we might learn that the speaker risks great financial loss if the individual in question follows through with her duty. In case 3 we might learn that the speaker himself has a tendency towards impulsive behavior, and tries his best to sympathize with those who are similar.

For every example of normative Moorean absurdity we raise, there is always the potential for a compelling explanation to be in the offing. An even more extreme embodiment of this dissonance, the figure of the amoralist, should be familiar to those who have come in contact with the internalist/externalist debate in moral psychology. This is the individual who says unflinchingly with Milton's Satan "Evil be thou my Good." But as cases 1–4 demonstrate we don't need to look to severe depression, personality disorders, or literary characters to find situations in which the denial of motivation or affection in accordance with a speakers normative commitments can be made intelligible to us. Does such an admission spell defeat for the Practicality thesis?

It does not. My explanation moves in three stages. The first asks us to examine our reactions to an "all else is equal" version of these cases. That is, would these speakers cause us puzzlement if we *knew* that there was no further story to be told? I believe that it would. In fact, I suggest that it is hard to imagine a speaker who simultaneously (a) is in a context

in which they are taken seriously, (b) uses normative language while feeling no appropriate motivation or affective reaction (and are self-aware of this fact), (c) understands fully well the words they are using and is not intending to deceive or mislead, and (d) is not in some mitigating circumstances that explain this lack of motivation/affection. Part of what makes such situation hard to envision and evaluate is that it involves a conversation in which one interlocutor is misusing language in a way that is radically uncooperative. The principle of charity is so central to our ability to converse with one another, that dreaming up such scenarios feels stilted and artificial. I think that such a result counts in favor of the Practicality thesis.

This still leaves us in the seemingly uncomfortable situation of allowing for agents whose normative beliefs and motivations sometimes come apart from one another, and admitting that such events are commonplace.[12] This brings us to the second stage of explanation. Recall the observation from the previous section, that our normative considerations may often tell against one another. In such cases, the motivational or affective force of certain commitments can be outweighed and sometimes even blotting out.

For example, suppose you believe that your friend is in bad epistemic shape. At the same time you care about him and know that he isn't really causing any harm, so you can't really bring yourself to blame him. Or suppose I believe that my partner would be in the wrong not to come clean on the bank clerk's error, but we really need the money so I can't wholeheartedly recommend giving it back. Notice that these situations involve pitting our

[12] Remember, part of what it is to remain consistent with Psychological Humeanism is to make peace with the possibility of such a phenomenon.

values against one another. Part of what it is to imagine circumstances that render the four cases above intelligible is to imagine *countervailing* normative considerations whose grip on the speakers *overcomes* the judgments they explicitly avow. Again, I believe this counts in favor of the Practicality thesis.

The last step is to recognize the fact that we haven't really gotten clear on what is meant by "wholeheartedness" in our statement of Practicality at the beginning of this section. I take that to be a task for each hybrid theory to discharge in the way it deems best. However, what we have learned is that in order for a theory endorsing Practicality to be plausible, it must be able to make sense of the way that we can often use normative language with less than full psychological harmony. It may be that the speakers in the cases we have been thinking about are not capable of putting forward their normative claims in *all* seriousness. But there is some seriousness to what they are saying, and knowing the rest of the story prevents us from accusing them of outright insincerity.

Let us take stock again: part of what it is for thought and talk to count as normative *just is* for its serious application to be expressive of motivational or affective states. We are only tempted to doubt this when normative considerations are set against each other. Once we consider these cases carefully, rather than counting against the proponent of Practicality they make her case stronger. However, it is important to keep in mind that in everyday life normative beliefs, motivations, and feelings can genuinely lack harmony—a point of contact between common sense and the theoretical commitments of those sympathetic to Psychological Humeanism.

I'd like to close by acknowledging that, traditionally, something like the Practicality thesis has often been objected to on the grounds that it cannot fully account for the

cognitive, truth-conditional behavior of normative language. Of course, it is true that Practicality on its own cannot capture the cognitive status of normative speech and thought—in fact, it has nothing to say on the matter. However, in order for this observation to become a criticism, one would have to make the further claim that Practicality is *inconsistent* with a theory that fully captures the cognitive element. Hence, like the objection brought up in Part II's discussion of cognitivism, this should be seen not as a challenge to the positive claim of Practicality, but rather to the possibility of a fully satisfying hybrid metanormative theory. Again, we will have to wait for later chapters to see whether such a challenge may be met.

Part IV: A Brief Outline of the Remaining Chapters

My goal in this dissertation is to offer a new hybrid metanormative theory. This theory is novel both in its approach and in what it is able to accomplish. Unlike previous attempts, my theory explains the hybrid quality of normative thought and language use by appealing to our best current understanding in speech act theory. This enables my view to place the cognitive and practical aspects of normative speech and deliberation on equal footing—something that hybrid theories were previously unable to do. By taking a close look at the phenomenon of inner speech, my theory is also the first to tackle head-on the connection between normative speech and thought. This enables it to say something informative about what we are doing when we think about norms, unlike previous theories

that have only analyzed normative social speech and gestured suggestively at normative thought.

In the next chapter I examine two of the most recent and prominent attempts to construct hybrid metanormative theories. These two theories are Stephen Finlay's "End-Relational Theory," and David Copp's "Realist Expressivism." As I describe each view, I point out various recognized costs and flaws along the way. At the end of each discussion I raise my own new objections to each view. What emerges at the end of Chapter 2 is a list of challenges and requirements that my own hybrid metanormative theory must address in order to succeed where others have failed.

In Chapter 3 I lay out the details of my positive hybrid metanormative theory in the arena of normative social speech. I begin by discussing some fundamental ideas in speech act theory and developing criteria for when we are justified in positing new, previously unrecognized types of speech acts. After outlining my theory of normative speech as constituting distinct and novel varieties of speech acts *in abstracto*, I apply it to the specific cases of moral discourse and aesthetic discourse to illustrate its ability to solve outstanding puzzles in those domains. Chapter 3 concludes by distinguishing my theory from some other views that might at first seem similar.

Chapter 4 addresses the phenomenon of inner speech. In order to trace the implications of my hybrid metanormative theory of normative social speech for normative thought, a better understanding of inner speech is needed. The chapter begins with a survey of some of the empirical research into inner speech. I then grapple with the problem of how to understand what it is to perform a speech act in the context of inner speech and lay out a more general theory of inner speech acts. The chapter concludes with a discussion of

normative inner speech in particular. For most adult humans, a great deal of the activity of normative deliberation takes place via the medium of inner speech. A better understanding of the pragmatics of normative inner speech acts thus casts light on what we are doing when we think about norms. The dissertation concludes in Chapter 5 with a summary of my arguments and some suggestions for further research.

CHAPTER 2

WHAT'S WRONG WITH OUR CURRENT HYBRID METANORMATIVE THEORIES?

Part I: Introduction

In this chapter I discuss two examples of hybrid metanormative theories: Stephen Finlay's End-Relational Theory and David Copp's Realist-Expressivism. I have chosen these two views because they are the most recent and most fully worked out exemplars of two ends of the hybrid theoretical spectrum. Stephen Finlay's view is an example of an approach that pairs a fully descriptive semantics of normative terms with a pragmatic explanation of their practicality, an explanation that depends on particular features of conversations in which normative terms are used. David Copp's view is an example of an approach that incorporates the practical element of normative terms into their semantic content. Because Copp also aims for compatibility with moral realism, his is one of the most fully hybrid of such "expressivist" style views. What I will show in this chapter is that these two recent, prominent, and promising approaches to hybrid metanormative

theory have difficulty fully embracing both sides of the cognitive/practical divide. As we will see, the virtues of each approach highlight the vices of the other.

If my own view can embrace the virtues of these two theories while avoiding their vices, this will constitute a strong argument in its favor. By the end of the chapter, two other lessons will emerge as well. The first is that it is no small achievement for a hybrid theory to remain neutral with respect to the semantics of normative terms. If this is something that my theory can do, that will give it an advantage over many rival views. The second is that the current literature on hybrid theories has failed to adequately appreciate that translating explanations of normative speech into explanations of normative thought is a significant challenge in its own right. I will tackle this challenge head-on in the fourth chapter.

This chapter begins in Part II with Stephen Finlay's End-Relational Theory. Because his pragmatic explanation of the practicality of normative language *depends* on his controversial semantic view, it is with his semantic view that we will start. After describing Finlay's semantics, pragmatics, and my objections to them, the chapter transitions to Copp's view in Part III. As its name implies, Copp's Realist-Expressivism is most easily understood as a foil to the anti-realist/cognitivist approach that is championed by Finlay. Once I have described Copp's view and detailed my objections to it, I close the chapter by offering a brief summary of the lessons learned. This gives a sense of the subsequent requirements that I should aim to satisfy as I begin to work through the details of my own theory in the next chapter.

Part II: Stephen Finlay's "End-Relational Theory"

Section 1: End-Relational Semantics

In his book *Confusion of Tongues* Stephen Finlay constructs a theory of normative language that aims at analytic reductions of the meaning of normative terms. Finlay's proposal is to reduce normative concepts to non-normative concepts, and thereby to reveal the non-normative metaphysical structure underlying normative properties. He takes success for this view to be measured in simplicity, adherence to linguistic and ethical intuitions, and theoretical conservatism. He also intends his theory to be fully "hybrid" insofar as he aims to pair a descriptivist semantics with pragmatic explanations of ethical discourse's motivational expression. An ambitious project indeed.

Finlay proceeds in laying out his semantic view by giving non-normative reductions of the three thin normative terms 'good,' 'ought,' and 'reason. In each case he offers a single unified semantics for all normative and non-normative uses of the term. To illustrate this, let's start with his analysis of the term 'good.' Finlay claims that 'good' is an incomplete predicate that expresses a relation between two propositions. That is to say, whenever we say something is "good" what we really mean is that it is "good *for*..." something else. Uses of the term in which this structure appears absent are actually cases of ellipsis. For example, when I say "chocolate is good," what I really mean is that chocolate is "good for…" *something*. In different situations I will finish this thought in

different ways—for instance, I might mean "chocolate is good *for gustatory pleasure*" or even "chocolate is good *for recovering from a dementor attack*."[13]

But what does this "good for" relation amount to? The intuitive answer is something akin to "promoting" or "being conducive to." Finlay's answer is that "good for" *means* "increases the probability of." This makes sense of the gradability of the predicate. That is, 'good' comes in degrees, including the comparative 'better' and the superlative 'best.' What these additional predicates mean, respectively, is "to offer a greater increase in probability than some other alternative" and "to offer the greatest increase in probability of all relevant options." Let's apply this to our examples above: in the first instance Finlay will translate my "chocolate is good" claim to mean "eating chocolate increases the probability that gustatory pleasure will occur," and in the second instance to "eating chocolate increases the probability that I will recover from a dementor attack."

Though Finlay brings to bear additional resources in response to possible objections and complex examples, this is the essential core of his analysis of 'good.' However, there are two challenging questions that remain. The first concerns how Finlay will handle uses of 'good' that appear to be explicitly unrelativized (e.g. "good simpliciter" or "good in itself"). The second has to do with how Finlay will explain normative disagreement: the apparent fact that we can agree about the relevant probabilities while disagreeing about what is "good." I'll put off discussion of these issues for now as they arise for the other two terms as well.

---

[13] "Harry took the chocolate but didn't eat it. 'What was that thing?' he asked Lupin. 'A dementor,' said Lupin, who was now giving chocolate to everyone else. 'One of the dementors of Azkaban.' Everyone stared at him. Professor Lupin crumpled up the empty chocolate wrapper and put it in his pocket. 'Eat,' he repeated. 'It'll help.'" Rowling (1999), pp. 84–85.

The second term that Finlay analyzes is 'ought.' He proposes that all uses of 'ought' may be understood as elliptical for instrumental means-ends uses. So 'ought' may be defined as describing a means-ends relationship between two events, where that relationship informs us that the means is the most reliable route to achieve the ends. By "most reliable" Finlay means "increases the probability of the ends occurring more than the other salient options." So when I say "you ought to eat the chocolate," what I really mean depends on the end I implicitly have in mind. So on one occasion what I might really mean is something like "of all salient options, you eating the chocolate is the event that most increases the probability of the occurrence of your gustatory pleasure."

A couple final notes about 'ought.' On Finlay's view functional uses (e.g. "knives ought to be sharp") are elliptical for instrumental uses (e.g. "in order that they cut well, it ought to be that knives are sharp"). Also, 'ought to do' style recommendations of actions (e.g. "John ought to save the child") are elliptical for 'ought to be' descriptions of events (e.g. "it ought to be the case that John saves the child"). Importantly though, as in the case of 'good,' the issues of apparent non-relational uses (e.g. "you ought *categorically* to do *x*—no matter your ends") and the possibility of normative disagreement are both pressing. Those issues will be addressed below.

The final term that Finlay analyzes is 'reason.' He sets aside as beyond the scope of his discussion two uses of the term: 'reason' used as a verb to refer to a mental activity, and 'Reason' used as a noun to refer to a mental faculty. However, he does intend to give a unifying semantics for all of the following uses: normative reasons for action, motivating reasons for action, justifying reasons, explanatory reasons, objective normative reasons, and subjective normative reasons. The most fundamental use of 'reason', to which Finlay

reduces all others is that of an explanatory reason. More precisely, 'reason' picks out a relation between two facts. This relation consists in an "explanation why," which on Finlay's view means "information that reveals *what makes it true that…"*

So when I say "the fact that breakfast at the hotel ends at 9am *is a reason* for you to be downstairs by 8:30" what that means is that "the fact that breakfast at the hotel ends at 9am is information that reveals what makes it true that it *would be good* for you to be downstairs by 8:30" or "…that it *ought to be the case* that you are downstairs by 8:30." Once we apply the End-Relational analyses of 'good' or 'ought' to these definitions, we have a fully unified and reductive definition of 'reason.' Again, this is going to depend on what end the speaker has in mind. In one case, for example what the speaker might really mean is: "the fact that breakfast at the hotel ends at 9am is information that reveals what makes it true that you being downstairs by 8:30 is the event that would, of all relevant alternatives, most increase the probability that you are able to eat breakfast."

How, on Finlay's view, can a reason (an explanation) be relative to a subject? The answer is simple: by being an explanation relative to the information a subject possesses. This means that the analysis of 'reason' must disambiguate between the subject *to whose information* the reason claim is relativized and the subject *about whom* the reason claim is being made. Though the two may often be the same person, they need not be. So then, "*r* is a reason for person *s* to perform action φ" means "*r* is information that reveals to s_1 why it is true that it would be good for it to be the case that s_2 φs" (where s_1 and s_2 could be identical). Note that in cases where two *different* people are being spoken of this definition is still ambiguous between being an explanation to s_1 of why it is good *for* s_1 that s_2 φs, and being an explanation to s_1 of why it is good *for* s_2 that s_2 φs. In the complete analysis Finlay

makes sure this is fully specified, but such a level of detail will not make a difference to the remainder of the discussion in this chapter.

Another epicycle of complexity involves explaining the practical connection between reasons and agents, in addition to the epistemic connection between reasons and subjects. For example, given appropriate background information, *that Terry stole Victor's wallet* might count as a good reason (explanation) for why it "ought to be" that Terry is apprehended (e.g. because it is the event most likely to bring about the end of punishing lawbreakers). So it's a reason *for* both Terry and Victor in that epistemic sense. However, it is also natural to say that Terry's having stolen Victor's wallet is a reason *for Victor* that Terry should be apprehended. But it seems strange to say that his having stolen Victor's wallet is a reason, *for Terry,* that he should be apprehended.[14]

To finish off the analysis, Finlay proposes that the End-Relational aim of the goodness or 'ought' claim that the reason explains be relativized to the ends (desires) of the agent for whom (and *to* whom) the reason/explanation is being given. I believe that Finlay is successful in resolving this challenge as well. However, the issues of non-relativized claims and normative disagreement remain to be dealt with. It is to Finlay's treatment of these two issues that we will now turn.

Section 2: Categoricity and Disagreement

Finlay's theory of normative language is intended to give unified, reductive analyses of all normative and non-normative uses of terms like 'good,' 'ought,' and

[14] Unless, of course, Terry lives in a society committed to a form of punishment, the subjection to which is guaranteed to make Terry himself better off in the long run. Such a scenario is, admittedly, far-fetched.

'reason.' His proposal is that our usage of these terms is always End-Relational, either explicitly or elliptically. However, in their most *characteristically normative* uses—especially in moral contexts—these terms seem to express explicitly non-relativized or "categorical" claims. When hedonists say that pleasure is good, they don't mean that it is good *for* something. They mean that it is good "in itself." When Kant says that you ought not lie, he does not mean that lying will decrease the probability of an outcome you already happen to care about.

Let's take categorical 'ought' claims first. When speakers make these claims, they seem to feature a kind of inescapability. Finlay's account is compatible with it being *true of* an agent that he ought to do something in order to achieve some end, even when he doesn't care about that end. For example, imagine Robert is hacking away at a large tree with a small hatchet. Suppose further that Robert doesn't care about successfully cutting the tree completely down, he only wants to vent some frustration. According to Finlay's view, it may still be *true of* Robert that in order to fell the tree he ought to use a chainsaw. Though unqualified normative claims are ordinarily interpreted as being relativized to the ends of the subject *to whom* or *about whom* the claim is being made, a speaker may relativize her claim to whatever ends she wishes.

So Finlay is able to make sense of the fact that speakers may truly claim of agents that they *ought* to do things that do not promote their (the agent's) ends. Therefore, if there is a sense of "inescapability" that *does* present a problem for Finlay it must be a pragmatic sense. Here is how he puts the challenge in premise conclusion form:

PC1. If an 'ought' claim is relativized to an end *e*, then addressing it to an agent *s* who lacks a preference for *e* doesn't promote conversational ends.

PC2. Any utterance that doesn't promote conversational ends is pragmatically inappropriate.

PC3. If an 'ought' claim is moral then no end *e* is such that if *s* lacks a preference for *e* then addressing the claim to *s* is pragmatically inappropriate.

PC4. Therefore if an 'ought' claim is moral, it isn't relativized to an end.[15]

In other words, Finlay clarifies that on his view a speaker can sometimes raise normative claims that are relative only to the speaker's own ends, and not at all relevant to ends her audience has. This explains why it can be *true* of an agent that he ought to do something even when he doesn't care about the end to which the 'ought' claim is relativized. According to Finlay, the remaining challenge is to explain why in some cases it makes sense, given conversational ends, for the speaker to keep insisting on her 'ought' claim even after she finds out her audience doesn't care about the end to which it is relativized. The moral case seems like one where this insistence *is* appropriate, but Finlay's view at least seems like one that would count all such cases as being *inappropriate*.

Finlay's solution is that the Instrumental Law of Pragmatics (always speak in such a way as to best achieve your own conversational ends as speaker) is more fundamental than the Cooperative Law of Pragmatics (always speak in such a way as to best achieve you and your audience's collective ends). Even though speakers *often* aim at cooperation, they *always* aim at their own ends in conversation. Sometimes these two goals can come apart. In moral cases, the speaker's ends are often more important to her than cooperation. This explains the insistence, and its appropriateness. But if the speaker's goal is to *promote* certain ends, why does she think that it is worthwhile to continue asserting her normative

[15] Finlay (2014), pp. 179–180.

claim after it has become clear that her audience is indifferent to her those ends? And why does it seem appropriate for the speaker to persist even when she realizes her audience already has the information that she is trying to communicate?

Finlay explains this persistence by appeal to the mechanism of *rhetorical objectivity*. Rhetorical objectivity is a phenomenon at play in situations where a listener can infer that the speaker is not trying to get the listener to *believe* an asserted proposition, so the listener will understand the speaker to be advancing a proposition with *non-assertoric force*. When it is in common ground that the listener can make happen the event for which the speaker is expressing a preference, the listener can interpret the speaker as putting forward a claim with imperative or prescriptive force. For example,

> "You don't want to do that."
>
> "The rubbish bin is over there."
>
> "Nobody in our family belches at dinner."

These indirect speech acts, interpreted as being put forward with imperative or prescriptive force, are much stronger than

> "I'd *prefer* you didn't do that."
>
> "I'd *like* you to put your rubbish in the bin."
>
> "I'd *rather* you didn't belch at dinner."

Pragmatically, expressing *intolerance of refusal* is a much more forceful way of prescribing than merely expressing preferences.[16] According to Finlay, when speakers use thin normative terms to make claims about ends *as if* those ends are shared by the audience, even when this presupposition is transparently false, a species of rhetorical objectivity called "moralism" is at play. Moralism is a rhetorical device—like sarcasm, hyperbole, or irony. It is a mechanism by which a speaker can use a declarative sentence to indirectly perform a non-assertive speech act like prescribing or demanding.

While moralism may explain some cases of categoricity, Finlay admits that his response ultimately just pushes the problem back. After all, don't speakers often take themselves to be appealing to a fundamental end that is preferred, but not because it promotes any further end? An ultimate or "final" end is an end that is naturally judged to be "good" or what "ought to be" *for its own sake*. Final value is a kind of value that is essentially "non-instrumental" or "intrinsic." In contrast, End-Relational value is value that is grounded in a relation to an end, and so might seem necessarily extrinsic.

To make matters worse, that which is "instrumentally valuable" is actually a proper subset of that which is "End-Relationally valuable." This is because on Finlay's view it is trivially true that *everything* is End-Relationally valuable (for something!).[17] But if an object's instrumental value is derived from its relation to an object with final, intrinsic value, then Finlay's theory cannot fully account for instrumental value either. Thus it looks

[16] At least, all else being equal. One can easily imagine a context in which the speaker has a history of extreme passive aggressiveness that causes the audience to interpret these latter three claims with as much (or even more) strength.

[17] To demonstrate this point Finlay gives "a simple proof by reductio: if anything were not good for some end or other, then it would be good for proving this claim false." See Finlay (2014), p. 198.

like Finlay can't accommodate either sort of value (instrumental or final), since they both bottom out in claims about what is good "full stop." To fill this gap, Finlay proposes to analyze "good for its own sake" in a way that is consistent with its being something that we desire, and not because it will get us some other thing. Remember, he is committed to saying that things are good because they increase the probability of ends. In a nutshell, his solution is just to say that *one really great way for some x to increase the probability of some y, is for x to be* ***identical*** *to y*.[18]

Thus, for example, to judge that "pleasure is good" is to judge of some saliently desired end y that pleasure is good for y. To say that pleasure has intrinsic value is to say that, by occurring, pleasure increases the probability of its own occurrence.[19] In keeping with our intuitions about intrinsic value, pleasure instantiates this property on account of its own nature—independently of anything else——in virtue of possessing the reflexive relational property of self-identity.

It should be clear by now that Finlay's account renders *all ends* intrinsically good. This is follows from the pedestrian fact that all ends are self-identical. So the problem of disagreement is now twofold: Finlay needs to explain both how two speakers who agree about the relevant probabilities could meaningfully disagree about a normative claim, and also how two speakers could meaningfully disagree about whether something is intrinsically valuable. It looks like all denials of intrinsic value should come out false on Finlay's view. Here's how he puts the problem:

---

[18] Finlay (2014), p. 200.

[19] Importantly, according to Finlay this would not also amount to *saying* that one desires pleasure, at least not in virtue of the semantic meaning of the statement you have asserted. However, under normal conditions such an assertion would also *pragmatically communicate* that one desires pleasure. More on Finlay's pragmatics in the next section.

> [S]uppose Saddam Hussein declares to his generals, 'We ought to use our chemical weapons on the Kurds'; surely we disagree morally with this claim if with any at all, yet our theory suggests that he may have asserted only a true, perhaps even obviously true proposition; similarly with many other appalling claims. Or suppose he continues, 'If the Kurds suffer agony as they die, this would be good … for its own sake!' Our analysis of final value suggests that the asserted proposition is roughly that *the Kurds' suffering agony would be good for the Kurds' suffering agony*, which is trivially true. We appropriately disagree with many such claims about final value, though Chapter 7's analysis seems to imply otherwise.[20]

A striking problem indeed. Finlay begins to address it by discussing the issue of informational relativity. Suppose two speakers make instrumental claims about what is good or ought to be in order to achieve some end, relative to the information available to each of them. Because their information is different, according to Finlay both claims might be true even though they appear to be contradictory at the surface level. In other words, the End-Relational view depends on a contextualist semantics. The *meanings* of normative terms are relativized to background information, just like indexicals. Consequently, in different contexts the same normative sentence can be used to express different propositions (just like the sentence "I am here" can be used to express different propositions in different contexts).

Finlay's view of normative disagreement essentially boils down to the following:

> *Robust Inconsistent Hypothetical Preference*: B's assertion that *p* normatively disagrees with A's assertion that *q* if B thereby indicates that if he were to prefer *e* he would on the basis of his information i_B have a preference inconsistent with the preference A thereby indicates she would have on the basis of her information i_A if she were to prefer *e*, and $i_B \geq i_A$.[21]

[20] Finlay (2014), p. 208.
[21] Finlay (2014), p. 228

According to Finlay genuine normative disagreement amounts to disagreement either in preference for final ends or in means/ends beliefs. In cases of informational disagreement, this may be plausible. However, as Finlay admits, this account is less satisfying when it comes to distinctively moral, categorical, or final uses of normative terms:

> Evaluative words like 'true', 'false', and 'knows' are used to express normative agreement and disagreement in preference, *as if* this were simply disagreement in belief over nonrelational propositions. This quasi-expressivist solution again has some significant advantages over expressivism proper. It doesn't require any radical revisions to naive semantics…It is fair to object that these "advantages" come at the cost of an error theory about some parts of the ordinary practice of evaluating and individuating normative claims or judgments, which is radical in rejecting the ordinary assumptions behind the practice.[22]

According to Finlay, moral disagreement always amounts to either informational disagreement (disagreement about means/ends reasoning) or disagreement of preference. To answer our two questions above: if two people agree about all the relevant probabilities then their normative disagreement can only be one of preference, and if two people agree about what each other's desires are then their normative judgments cannot logically contradict one another. Finlay takes this resulting relativistic error theory to be a virtue of the End-Relational view: morality, as a practice, depends on its participants systematically denying the relativistic semantics of the terms they use. This "moralistic" use of rhetorical objectivity is what gives moral claims their prescriptive force.

Finlay does admit that his view leads to a quasi-self-effacing sort of discomfort:

> [I]t strains credulity (and moral decency) to claim that by saying, reprehensibly, 'We ought to use our chemical weapons on the Kurds', Saddam Hussein might actually say something true….since moral ends are

[22] Finlay (2014), p. 242, emphasis added.

> expected to be preferred even to philosophical ends of objective truth. If I am asked, 'So, do you think that Hussein said something true?', I feel trapped in a dilemma. To say that I do is, by virtue of moral pragmatics, infelicitously to indicate that I lack a certain "warmth in the cause of virtue". By "what he said/believed", we commonly think of the overall mental state expressed thereby, and evaluating it as "true" is naturally interpreted as expressing agreement in attitude….pressuring me to assent to the question 'Do you think what he said/believes is true?' is from this perspective just a rhetorical ploy aimed at morally shaming me into disowning what I believe to be the philosophical truth.[23]

My aim is not to shame Finlay, so I will leave the anti-relativistic arguments for others to make.[24] I am more interested in Finlay's moral pragmatics and whether, even setting aside the palatability of the semantics on which they depend, they are able to accommodate the target phenomena. It is to Finlay's handling of this issue that we will now turn.

Section 3: Moral Pragmatics

The pieces are in place. Finlay has proposed a thoroughly unified, reductive, descriptive semantics of normative terms—at the costs of siding himself with error theory and moral relativism. If his End-Relational theory doesn't have the resources to explain the *practicality* of normative language and thought, at least as well as expressivist rivals can, then it will all be in vain. Finlay claims he can answer the challenge by appealing only to one foundational pragmatic principle and a minimally Humean theory of mind.

Finlay's first step is to posit a robust distinction between semantics and pragmatics. Semantic information is encoded in the sentence, while pragmatic information is generated by the act of uttering a sentence in a certain context. So pragmatic information is a function

---

[23] Finlay (2014), p. 244.

[24] Though it is worth asking whether Finlay is attempting his own rhetorical ploy to trap his *critics* in a dilemma.

of the semantic information taken together with contextual information. One important piece of the context is *common ground*, the background assumptions mutually understood by speaker and audience *to be* mutually understood.

The distinction between semantics and pragmatics can be seen most clearly in cases of irony and sarcasm. Suppose we are looking through your beautiful vacation pictures together and—upon arriving at an artistic capturing of a particularly blissful moment on the beach—I remark, "This looks like a really miserable trip." The sentence I have uttered expresses, in virtue of its grammatical structure and the meanings of the terms I have used, that your trip appears to have been an extremely unenjoyable one. Of course, given the context we are in, my use of a sentence like that with just the right inflection and facial expressions is able to convey, pragmatically, that your trip appears to have been amazing.

In this example, what I have *said* is that your trip looks miserable. What I have *meant* in saying what I did is that you trip looks wonderful. Sarcasm is a tool that enables us to *mean* something other than what we have *said*. But this is just one particularly vivid illustration of the distinction between sentence meaning and speaker meaning. On the traditional view, speaker meaning refers to what a speaker intends to convey to an audience, at least partly in virtue of their recognizing her intention to convey it.[25] In saying what I did I intended to convey to you that your trip looks wonderful, and I intended to convey that to you at least partially by getting you to recognize that this was my intention. However, by saying what I did I may have communicated even more than I have speaker-

[25] This is a simplified version of Grice's definition, however it quickly runs into difficulties. See Green (2007) pp. 53–63. For one thing, it does not allow for the phenomenon of speaker meaning when there is no audience present. Extended discussions of speaker meaning will be offered in the next two chapters.

meant. That is, you might be able to infer from my utterance that I am also jealous of you. This is additional pragmatic information, though it goes beyond speaker meaning. It could be communicated to you either intentionally or unintentionally.[26]

It is surprisingly difficult to draw a hard and fast line within the category of pragmatic information between what is and is not a part of speaker meaning. Finlay makes no such commitment, as he wishes to include in his discussion both what is and what is not intentionally communicated. His primary purpose is to achieve his explanatory goals while avoiding the charge of being ad hoc. The best way of doing so, Finlay thinks, is by "identifying a minimal set of independently motivated and compelling fundamental principles of pragmatics, and then deriving all our solutions directly and systematically by applying these to our semantic theory."[27]

To begin with, here's a principle that seems obviously true:

> *Instrumental Principle of Conversation*: In order to achieve your conversational ends you ought to speak in a way that is best for those ends.[28]

And it supports the following foundational principle of pragmatics:

---

[26] In one context, perhaps, I am trying to *hide* my jealousy under the guise of sarcastic light-heartedness, but you know me well and see through the ruse. In such a case the information that I am jealous would be communicated unintentionally. In another context, I might be trying to *express* my good-natured jealousy by giving you a hard time. In that case I would be intending to communicate the information to you. In both cases you are able to pick up on this information via contextual clues, even though it goes beyond both what is said and what is speaker-meant.

[27] Finlay (2014), p. 121.

[28] Note that if one accepts Finlay's reductive analysis of 'ought' this principle comes out analytically true: "in order to achieve your conversational ends, your speaking in the way that most increases the probability of your achieving your ends is the way of speaking that would, of all relevant alternative ways you could speak, most increase the probability that you are able to achieve your ends."

> *Instrumental Law of Pragmatics* (ILP): Speakers always speak in the way they believe best for their conversational ends.

As noted above, Finlay takes ILP to be more fundamental than Grice's Cooperative Law of Pragmatics. After all, Grice's cooperative principle (that speakers always speak in such a way as to best achieve their and their audience's collective ends) isn't meant as a universal generalization of how speakers actually behave, just an important presumption that usually works (given that speakers and audiences often have shared conversational ends). Finlay argues that by combining ILP with his semantic theory, he can give a satisfying pragmatic explanation of the practicality of normative language.

Finlay boils the practicality challenge down to three main desiderata:

> (a) Motivational Internalism: any agent who judges himself to have a reason to φ must have some motivation to φ.
>
> (b) Interpersonal influence: normative utterances characteristically have motivational effects on others—getting others to share your normative judgments will cause them to be similarly motivated.
>
> (c) Practical Illocutionary Force: normative sentences are used to perform speech acts of distinctly practical kinds (endorsing, recommending, prescribing).[29]

The problem for a descriptive semantics is that normative terms are said to only contribute to the descriptive content of sentences. Sincere assertions of such sentences require only beliefs—no motivation, desire, or intention. Of course, descriptivist views can pragmatically explain various instances in which descriptive sentences are used practically, given the right context—especially the right common ground. If it is mutually understood

[29] Finlay (2014), p. 126.

that your friend appreciates sunny houses, you can *recommend* a house to her merely by describing it as one that gets a lot of sunlight.

The real problem is to explain the *especially close* connection between practicality and characteristic uses of normative sentences, given a descriptive semantics. Finlay puts the purported inconsistency like this:

> (1) Strong Practicality Requirement: sincere normative utterances entail some degree of speaker motivation, and necessarily involve some practical kind of force such as endorsement or recommendation.
>
> (2) Humeanism: beliefs and desires are fundamentally different types of mental states (different directions of fit), and they aren't ever necessarily connected to each other (e.g. no belief always entails motivation).
>
> (3) Descriptivism: normative sentences have purely descriptive content, so their sincere assertion only necessarily expresses beliefs, and can express desires or motivation only contingently.[30]

Expressivists conclude that the only way to adequately account for motivational internalism is by encoding the practicality in the semantics, so they deny (3). Finlay proposes to "beard the lion" by holding on to both Descriptivism and Humeanism. He does this by modifying the practicality requirement. This is ok, he says, because it is almost universally agreed that the strong version is *too* strong.[31]

To show this, Finlay first points out that the three normative terms he has chosen to analyze have both normative and non-normative uses that don't seem to imply motivation. For example, if I say "Adultery is good for destroying marriages" I do not thereby imply anything about my own motivational states (unless information about my

[30] Finlay (2014), pp. 128–129.
[31] Finlay (2014), p. 130.

opinion of the value of marriage is already in common ground). Or suppose I say "the tree has good roots." Such a statement on its own does not convey to you any of my motivational or affective states. An even bigger problem for the strong practicality requirement, however, is the issue of embedding:

> Logically simple normative sentences generally express motivation and have practical illocutionary force only when directly asserted, and lack these features in almost every other kind of use, as when embedded in negations, disjunctions, conditionals, attitude reports, and questions. Whereas saying, 'You ought to φ' is plausibly to perform a speech act of recommendation and implies some kind of corresponding speaker motivation.[32]

Finlay notes that, besides creating problems for expressivists, this feature of normative language creates problems in general "for the view that normative words have their illocutionary force of practicality by semantic convention."[33] This observation will resurface in our discussion of David Copp's view below.

Here are two weaker principles of practicality that Finlay accepts:

> *Indispensability*: Any agent, whatever her desires and intentions, necessarily employs concepts that actually can be expressed by the words 'good', 'ought', and 'reason' in her practical thought aimed at decision and action.
>
> *Evidentiality*: Assertion of an unrelativized normative sentence is sufficient though defeasible evidence of corresponding motivation, independently of any information about the speaker's desires.

Indispensability may be easily accommodated by the End-Relational theory with no additional pragmatic explanation needed. Given that agents desire some ends, things that

---

[32] Finlay (2014), p. 131.
[33] Finlay (2014), p. 132.

make those ends more probable will serve a practical role for the agent. *Evidentiality* is trickier. According to *Evidentiality*, when a speaker leaves out mention of the ends to which her normative claim is relativized—as is the case in moral contexts—the speaker communicates that she is motivated. Finlay is committed to these "unrelativized claims" always being elliptical for relativized ones, so he has to explain why it is precisely *the speaker's making implicit* of the relativization that causes her motivation to be expressed.

The Cooperation Law of Pragmatics requires that a speaker only use ellipsis when the element she is leaving out is salient enough to be easily identified by her audience in the context. In a given conversation, the most salient end will be the speaker's *conversational end.* According to Finlay, this means that when the end of a normative claim is left implicit, looking to the speaker's conversational ends and her own personal desires will usually enable the audience to determine what her statement is elliptical for.

On Finlay's view, when a speaker says that "x is good" she is saying that x raises the probability of some end she desires. Hence it can be inferred that she desires x (the means to the end) as well. In this way she expresses approval of x. Given that her utterance is likely to indicate her approval, the audience is justified (ceteris paribus) in understanding her to have spoken with the communicative intention of expressing this approval. Alternatively, the speaker's normative claim could be interpreted as being about x's raising the probability of something her audience desires. In such a case, the speaker can be understood as seeking to influence the motivations of her audience.

Finlay takes his view to have several advantages over both expressivists and primitivists. Remember, expressivists analyze normative terms as semantically expressing a speaker's motivational or affective states. Primitivists, on the other hand, deny that

normative terms may be reductively analyzed. The first advantage Finlay sees his analysis of the practicality of normative language as having is its nice synergy with his unified semantics for 'good', 'ought', and 'reason'. He also explains practicality in a way that is consistent with Humeanism—unlike many primitivist views, he requires no radical claims about human psychology, motivation, or agency. Moreover, his theory enjoys over the expressivist all the benefits of a straightforward traditional semantics (without the non-reductive commitments of the primitivist). According to Finlay, normative claims are just disguised claims about non-normative objective truths—disguised in a way that gives them greater rhetorical power.

Finlay also claims his view has advantages over some other hybrid theories—namely those like David Copp's that tie the expression of motivation to semantic conventions. This is because his view can explain non-practical uses of normative language (e.g. when normative predicates are embedded in conditionals—"if eating meat is wrong then eating this hamburger is wrong"—or explicitly relativized—"sharp knives are *good for* cutting"). In contrast, Finlay describes his view as turning on "contingent features of the context of use."[34] He cites Grice on the issue of cancellability, admitting that on his view the motivation expressed by a speaker making a normative claim may be cancelled *in the Gricean sense*.[35] I will address this issue in greater detail in the next section, as I believe it creates a significant problem. Finlay's appeal to conversational cues rather than conventional ones is consistent with his earlier, conversational implicature approach.[36] Really the only difference is that his view is now married to a more fully developed

[34] Finlay (2014), p. 144.
[35] Finlay (2014), p. 144.
[36] See Finlay (2004) and (2005).

semantic theory. As we will see, this means that his theory is still subject to the same weaknesses.

Section 4: Objections to Finlay's View

The most natural cluster of objections to raise for Finlay's view take aim at the fact that it depends entirely on his End-Relational semantics. This means that his pragmatic solution is quite narrow, along multiple dimensions. For starters, notice that Finlay takes his view to be fully *metanormative* in the sense that he is offering an account of all normative language. However, he only explicitly discusses the three thin normative terms 'good,' 'ought,' and 'reason.' We are left to accept on faith that similar analyses may be given for all other normative terms (both thick and thin). Any skepticism about Finlay's ability to offer a satisfactory account of thick normative terms will count against the unencumbered metanormativity of his solution. I leave aside pressing this particular issue, except to note that if my theory can *demonstrate* an ability to analyze a wider range of normative terms, then it will issue a verdict where the incompleteness of Finlay's solution in an important sense still leaves the jury out.

There are two further paths down which objections might be directed specifically against Finlay's dependence on the End-Relational semantics. The first has to do with the relational analysis itself. For example, any independent arguments for thinking that normative terms are sometimes used in unrelativized ways (without being elliptical) will constitute objections to the End-Relational theory. Additionally, the commonsense intuition that speakers who make claims about what is intrinsically valuable are not merely making trivial claims about self-identity (or even about their own preferences) counts

against the End-Relational theory. Finally, any considerations in favor of the idea that normative terms do not admit of non-normative definitions—whether or not normative properties are metaphysically reducible—cast doubt on the End-Relational theory.

The second ground for dispute is the End-Relational theory's implication of moral relativism. Here is Finlay on the upshot of the End-Relational theory for normative ethical inquiry:

> The philosophical hope that normative ethics might be a science, like mathematics, is in vain. Although there is plenty for philosophers to say about morality and moral issues (clarifying distinctions, identifying inconsistencies, drawing implications), to make either a moral or an all-things-considered normative claim, like 'One ought always to maximize happiness', is fundamentally to express a contingent and subjective preference. Claims of basic moral principles are little more than coercive expressions of preference….
>
> The grand philosophical ambition to resolve normative disagreement by the use of reason alone would seem futile. This could turn out to be unduly pessimistic, of course. Perhaps human nature is at base sufficiently homogeneous that if all distinctions were clarified, all inconsistencies identified, all implications drawn, and all factual errors corrected, we'd find convergence in moral perspective, as philosophers have often hoped. This wouldn't make our moral claims any less perspectival, but would diminish the influence of factors beyond the jurisdiction of philosophy or science. However, to me as to many others this looks like wishful thinking, and so I believe that the End-Relational theory justifies skepticism toward normative ethical theory about "first principles".[37]

Regardless of whether Finlay's skepticism is correct, it is important to keep in mind what it would take for it to be *earned*. Finlay offers no normative ethical arguments for his position. Rather it is a byproduct of his semantic analysis of three thin normative terms, an analysis which in turn is the result of reading off of—and into—the surface grammar of

[37] Finlay (2014), pp. 256–257.

common ways of speaking, with the goal of providing a singular, unified meaning across all normative and non-normative uses of those terms. This analysis also plays an essential role in his pragmatic explanation of the practicality of normative terms. Thus any independent argument against moral relativism will constitute an objection both to the End-Relational theory in general, and to Finlay's pragmatic explanation in particular.

Any view that can achieve the same explanatory power, without opening itself up to these lines of objection, will have a significant advantage over Finlay's view. With that said, my goal for my own theory is to end up being compatible with *any* broadly descriptivist semantics, Finlay's included. Hence I will not (and need not!) pursue further any of these possible criticisms of the End-Relational theory in particular. A more difficult and interesting task is to come up with reasons that even a *dyed-in-the-wool supporter* of the End-Relational theory should have for rejecting Finlay's explanation of practicality in terms of conversational pragmatics.

The central problem of Finlay's explanation is one that it inherits from his older conversational implicature approach. Because Finlay explains the practical side of normative language by appeal to the "contingent features of the context of use" he is fully committed to the position that the expression of motivation by normative judges is cancellable in the *Gricean* sense.[38] Certainly there are situations where normative terms can be used without expressing practical attitudes (this will resurface as a problem for David Copp's view below). However, I submit (contra Finlay) that when normative terms *are* used in such a way as to convey a practical force, that expression is (in Grice's sense) *noncancellable*. To see this we need a clear statement of what it would take for the

[38] Finlay (2014), pp. 144.

phenomenon in question to satisfy Gricean Cancellability, which it will be helpful to distinguish from another sense of cancellability which I will term "In Principle Cancellability."

> *In Principle Cancellability* implies that, though an audience might at first be puzzled by a speaker's outright denial of something implicated by her utterance, there is some explanation she could possibly go one to give such that the audience's puzzlement would be eased.

When it comes to the expression of motivation by speakers who use normative terms, *In Principle Cancellability* seems extremely plausible. Suppose the speaker says at first "capital punishment is wrong" and then immediately follows this up by saying "I am completely indifferent to the thought of a criminal being executed." The audience would at first (rightly) be puzzled. Typically, when a speaker claims that something is wrong she thereby implicates that she feels some disapproval towards it. However, there are explanations that she could offer that would ease the audience's puzzlement. The speaker might clarify that her use of normative language is a non-normative one, an End-Relational one, or is in "inverted commas." She might describe her own position as a someone who is suffering from a personality disorder or severe depression. She might explain that her bare cognitive assent to the normative claim was obtained by a seemingly sound deductive argument, or that her feelings are simply overwhelmed by other normative considerations (a phenomenon discussed in the last chapter). The speaker could even dispel the audience's confusion by simply stating that she has suddenly changed her mind. Commitment to *In Principle Cancellability* just amounts to recognizing that a speaker can make a normative claim in a context where it is received in an ordinary, serious way, and then go on to cancel

her expression of motivation or affect by explaining either that she did *not* mean it seriously or that the situation *is* abnormal in some way.

In contrast, here is what is ordinarily meant by pragmaticists when they speak of cancelability:

> *Gricean Cancellability* implies that a speaker can make a statement that generates an implicature, explicitly deny what was implicated with no further explanation whatsoever, and thereby create no confusion in her audience.

An uncontroversial example of the phenomenon of Gricean cancellability will be instructive here. Suppose I say "I broke a finger this morning." In most contexts there is a presumption at play that people who say "a finger" are speaking of one of their own fingers, hence as long as we are in such a context my statement will express that the finger I broke was my own. However, I can follow up my assertion with a simple denial of the implicated content (e.g. "It wasn't my finger that I broke") and the audience will effortlessly shift to interpreting my first claim as being about what I did to *someone else's* finger. This may prompt them to ask for more information (after all, it sounds like there must be an interesting story to tell) but even without further information they will not be left feeling confused. There will be no lingering sense that language has been used incorrectly.

In the case at hand, commitment to *Gricean Cancellability* amounts to claiming that a speaker can make a normative claim in a context where it is taken at face value, follow up the normative claim with an explicit denial of motivation or affect, offer no explanation whatsoever, and still leave the audience with no confusion or lingering feeling that language has been used improperly. As noted above, this is the sort of cancelability to which Finlay has committed himself. According to Finlay, when a speaker's use of a

normative term expresses motivation, it does so because the speaker leaves implicit the end to which the claim is relativized, *and* the context is one in which there is a general presumption that the speaker's claims are relativized to ends she desires. If Finlay is right, then all a speaker need do in such a context is add an explicit denial of motivation and the audience will effortlessly shift to interpreting the claim as being relativized to an end about which the speaker doesn't care. This may prompt them to ask for more information, but even without more information they should not (according to Finlay) be left feeling puzzled, misled, or with the lingering sense that language has been used incorrectly.

David Copp nicely summarized the problem with Finlay's view in a passage quoted in the last chapter. Here it is again:

> Anna says that capital punishment is morally wrong, yet whenever she hears that a criminal is about to be executed, she is indifferent. There is never any indication that she disapproves of capital punishment other than the fact that she says it is wrong. She would in fact concede that she has no negative attitude toward it, nor has she any policy that would lead her to oppose it. Something is amiss. If Anna were to claim that capital punishment is morally wrong but then add that she has no negative attitude toward it, we would be puzzled.[39]

Finlay's only response to Copp is to say that they merely have a "basic clash of intuitions."[40] However, this is highly misleading. In explaining why his intuitions are different from Copp's, Finlay appeals to the fact that normative language can *sometimes* be used in a way that does not convey motivational force. But this is implied by *In Principle*

[39] Copp (2014), pp. 52–53. For additional insightful criticism of cancellability from a somewhat different direction, see Cholbi (2009).

[40] Finlay (2014), p. 144.

Cancellability—and these are phenomena that Copp readily admits.[41] Hence Finlay's response to Copp does nothing to address the true objection. Finlay's view commits him to the stronger form of cancellability, *Gricean Cancellability*, and it is this commitment that does violence to our linguistic intuitions. There is no response available to Finlay that can explain them away, he can only bite the bullet.

The final obstacle facing Finlay's view he expresses most clearly himself:

> Despite the breadth of this conception of pragmatics, some will object that it's still too narrow to be the key to solving metaethical puzzles, because whereas the domain of pragmatics is speech or utterances, the challenges cataloged above arise also in normative *thought* or judgments. Pragmatic solutions are complained to be insufficiently general because they presuppose a conversation; for example, many of our results will turn on the significance of ellipsis.[42]

This is a significant challenge. Finlay responds by saying that "these objections overlook important connections and parallels between speech and thought."[43] He goes on:

> On one hand, insofar as judgments are themselves linguistically formulated, they are plausibly internalizations of speech ("speaking to ourselves"), and therefore will resemble contributions to imagined conversations and inherit many pragmatically influenced features. It's implausible that although when speaking aloud we may omit various words the audience can easily recover, when saying things to ourselves we must mentally token only fully explicit sentences: introspection finds similar patterns of ellipsis in thought as in speech. Devices like sarcasm, metaphor, and hyperbole—though their standard explanations are pragmatic—are found as readily in thought (e.g., sarcastically thinking, 'You're a really smart guy!').[44]

---

[41] Whether Copp has the resources to explain all these non-motivational uses of normative terms is another matter. However he does *wish* to accommodate this data. See Copp (2014) pp. 54 and 59.

[42] Finlay (2014), p. 120.

[43] Finlay (2014), p. 120.

[44] Finlay (2014), pp. 121–122.

I think Finlay is on the right track here. However, he has nothing more to say. The reader is left with a promise that elides over all the potential difficulties and complexities of describing the relationship between public and inner speech. Finlay seems to assume that there is little more to the challenge than saying the same thing in "two different ways."[45] Consequently Finlay does not consider, much less rule out, the possibility that the correct view of normative thought is *might well be inconsistent* with his treatment of practicality in the End-Relational view. For those who are most interested in analyzing moral judgment in the sphere of thought and deliberation, Finlay's theory falls short.

The connection between normative speech and normative thought is one that is of primary significance and fraught with difficulty. Pursuing the strategy of explanation via the phenomenon of silent speech requires a careful examination of research in empirical psychology, as well as a better understanding of speaker meaning and speech act theory. It is just this strategy I propose to pursue in later chapters in order to demonstrate that my own theories of normative speech and thought are mutually supporting.

Part III: David Copp's "Realist-Expressivism"

Section 1: Background

Finlay's pragmatic account is unattractive for two primary reasons. The first is its dependence on a controversial semantic theory. This leads to concerns about narrowness

[45] Finlay (2014), p. 121.

and moral relativism. The second is its explanation of the practicality of normative speech via "contingent features of the context of use." This causes it to run afoul of our linguistic intuitions regarding Gricean Cancellability. Having encountered these problems, David Copp's Realist-Expressivism is a natural place to turn for solutions. This is because, in the first case, he intends to construct a theory that is semantically conservative—that is compatible with a wide array of descriptive semantic analyses of normative terms. Secondly, Copp takes the opposite approach to Finlay in analyzing the connection between normative terms and the expression of practicality: rather than appealing to conversational pragmatics, he takes practicality to be a component of a normative term's *semantic expression*.

In 2001 Copp published a paper titled "Realist-Expressivism: A Neglected Option for Moral Realism." In that paper, his goal was to propose a theory of ethical speech and thought that would show that realist ethical naturalism could be consistent with a robust form of expressivism. The resulting view was both cognitivist and a close cousin of internalism. I say a "close cousin" because Copp (along with some others[46]) takes motivational moral judgment internalism, at least in its most orthodox form, to be the thesis that having a belief with moral content necessarily implies being motivated. Internalism of this stripe is incompatible with Psychological Humeanism. Hence Copp and many other hybrid theorists reject it for the reasons I laid out in the previous chapter.

In broad strokes, Copp's original view was that moral predicates have (at least) two semantic roles. Their chief semantic role is to refer to moral properties, while at least one of their secondary semantic roles is to express motivational states of mind. This expression

[46] Schroeder (2009).

Copp termed "Frege expression" as his favored analysis of the phenomenon was inspired by the Fregean theory of "coloring." Of course, Copp admitted, such a view would be "ad hoc and unhelpful" unless it could avoid "postulat[ing] the existence of otherwise unexpected linguistic conventions."[47]

Copp sought to discharge this theoretical burden by subsuming Frege expression (including both the expression of negative attitudes by pejoratives and slurs, and the expression of motivational states by moral predicates) under the general category of Gricean conventional implicature. Conventional implicature is a mechanism by which speakers may express something in virtue of the "linguistic conventions governing the uses of the words in question."[48] Conventional implicatures are distinctive in that they do not contribute to the truth conditions of what a speaker has explicitly said, are detachable, and are noncancellable. An example here will be helpful.

Consider the following:

(1) Sally is a Canadian, and is therefore friendly.

(2) Sally is a Canadian and Sally is friendly.

(3) Sally's friendliness *is implied by* her being a Canadian.

(4) I don't mean to suggest that Sally's friendliness *is implied by* her being a Canadian.

According to the traditional view of conventional implicature, statements (1) and (2) have the same truth conditions—they are true just in case Sally is both a Canadian and friendly, and false otherwise. However, a speaker who asserts (1) also thereby *expresses* (3). This

[47] Copp (2001) p. 175.
[48] Neale (1999), p. 53.

expression is conveyed in virtue of the conventional meaning of ‘therefore’ (or as Copp would say, it is part of the Fregean “color” of ‘therefore’). This can be shown by observing that the speaker’s assertion of (1) also passes the Gricean detachability test: the speaker *could* express her belief in (1)’s truth conditional content *without* also expressing the further content of (3). The speaker could easily do this by asserting (2) instead of (1). Additionally, we can see that (1) passes the noncancellability test: a speaker who followed up her statement with an explicit denial of the implicated content would create confusion in her audience. Indeed, if our speaker asserted (1) and (4) in succession, without any further explanation, we would be left with the sense that she had used language incorrectly.

Copp took the expressions generated by the Fregean color of terms—“Frege expression”—to be an instance of conventional implicature because it does not contribute to the truth conditions of utterances, it passes the detachability test, and it passes the noncancellability tests. Among other things, according to Copp, Frege expression will include slang terms like ‘cur,’ pejoratives like ‘jerk,’ epithets like ‘damn,’ slurs like ‘dago,’ and moral terms like ‘wrong.’ Hence, Copp claims, his theory does not “postulate the existence of otherwise unexpected linguistic conventions,” rather it just appeals to linguistic conventions that we already recognize as being in place.

In the normative case, the result is a theory according to which asserting a moral claim like “torture is wrong” primarily expresses the truth apt proposition that torture is wrong as well as the speaker’s belief that torture is wrong. At the same time, due to the color of the term ‘wrong’ (part of its conventional meaning) a conventional implicature is generated to the effect that the speaker has a practical disposition to accept an authoritative

standard that, in this case, prohibits torture. This is an elegant solution indeed. Unfortunately, unforeseen problems arose for Copp's analysis.

Section 2: The Move to "Simplicature"

In 2009 Copp published a revision of his view in the paper "Realist-Expressivism and Conventional Implicature." This revision came in response to his becoming convinced (mistakenly I think) that Gricean implicatures are always a part of speaker meaning, while Frege expression is not always an aspect of speaker meaning.[49] He also used the opportunity to clarify why his revised conventional *simplicature* view was not subject to criticisms that have cast doubt on the existence of the phenomenon of conventional implicature in general.[50]

As we consider the further developments of his view, it is important to keep in mind the theoretical burden that Copp set for himself to discharge at the outset. He concedes that his analysis of moral concepts can only avoid "being ad hoc and unhelpful" if it confines itself to positing the same sorts of linguistic conventions that are (or should) *already* be doing good explanatory work for us. (This particular goal is modeled admirably in Finlay's view above, where we see him appealing only to the highly plausible Instrumental Law of Pragmatics.) If conventional implicature is a mechanism about which there is a great deal of doubt, and doesn't fit very well with the phenomenon of "Frege expression" anyways, then for Copp the burden remains.

[49] Copp (2009), pp. 175–177.
[50] For the most prominent of these, see Bach (1999).

The goal of Copp's 2009 paper was to discharge this responsibility by showing that conventional *simplicature* is a phenomenon with which we are already familiar, and which can continue to do good work for us. According to Copp, conventional simplicatures are detachable, cancellable, and do not contribute to the truth conditions of utterances. So far, so much like conventional implicature. The main difference is that conventional simplicature is a broader category—one that includes everything that a speaker communicates in virtue of the conventional meanings of the terms she uses, whether intentional or unintentional, as long as it is not part of what is explicitly said. More precisely:

> A speaker *conventionally simplicates* that *p* in assertorically uttering a sentence just in case
>
> *(a)* in assertorically uttering the sentence, the speaker communicates that *p* (whether intentionally or not), and
> *(b)* the fact that the speaker thereby communicates the proposition *p* is determined by the (or a) conventional meaning of some particular linguistic device in the sentence, but
> *(c)* the proposition that *p* is not part of what is said explicitly by the speaker in uttering the sentence, so that
> *(d)* the falsity of p is compatible with the truth of what is said explicitly by the speaker.[51]

So defined, Copp intends conventional simplicature to take over the same role that conventional implicature played in his original theory. All instances of Frege expression—slurs, pejoratives, moral terms, informal pronouns (like the French *tu* and the German *du*), and other terms with Fregean coloring—are instances of conventional simplicature. Hence, according to Copp, conventional simplicature is a familiar phenomenon that does a great

[51] Copp (2009), p. 184.

deal of work for us. Because it is not part of speaker meaning simplicature isn't a part of what is said, hence it can be shown to escape the problems that philosophers of language like Kent Bach have raised for the coherence of the notion of conventional implicature.[52] It looks like Realist-Expressivism is in good shape. However, Copp recognizes that there are two further potential problems that he needs to address.

Section 3: Response to Schroeder

In his 2014 paper "Can a Hybrid Theory Have It Both Ways? Moral Thought, Open Questions, and Moral Motivation" Copp's primary aim was to respond to two objections raised by Mark Schroeder against the possibility of an adequate hybrid theory. A secondary aim is to assuage concerns about the apparent disanalogies between different sorts of Frege expression. I believe that a closer look at the latter will show us that Realist-Expressivism runs into trouble, so I'll leave discussion of it off to the next section. For now, Schroeder's objections are aimed at all hybrid theories, so some insight will be gained from looking at how Copp responds to them.

Schroeder's first objection to hybrid views "having it both ways" is that in order to gain all the advantages of cognitivism they must be able to endorse the Inference Licensing Thesis. According to this thesis, valid arguments are such that accepting their premises commits someone, in some sense, to accepting their conclusion. For Schroeder "acceptance of a claim" amounts to being in the states one would express were one to sincerely assert the claim. The problem is that, according to a hybrid theorist, the fully sincere assertion of

[52] Copp (2009) pp. 180–182.

a moral claim expresses both belief and motivation.[53] However, it seems clearly possible that one could rationally accept the premises of a valid argument without being motivated in accordance with the conclusion.

For example:

P1. The suffering of animals is morally considerable.

P2. If the suffering of animals is morally considerable, then it is always wrong to consume animal products.

P3. If it is always wrong to consume animal products, it is wrong to sit on this leather chair.

C. It is wrong to sit on this leather chair.

Conceivably, one could believe all of these premises (and the conclusion) without being in all the psychological states that Copp thinks are expressed by a sincere assertion of the conclusion. Perhaps it was standing room only when you arrived late to Peter Singer's lecture. His arguments were compelling, but it's been a long hard day on your feet. By the time you arrive home, you'd love nothing more than to sink into your armchair and zone out in front of the television. In fact, you feel no disinclination whatsoever towards sitting in your chair. This certainly doesn't seem irrational, at least not in the way that we ordinarily associate with failing to appreciate a valid argument. How can a hybrid theorist like Copp maintain that the argument is valid, without at the same time ruling you as being irrational?

[53] For reasons to be explained in the next chapter, I believe this statement is actually incorrect. But until we have the resources in front of us to see exactly why, it will be better to speak *as though* this is what hybrid theorists must be committed to.

It turns out to be rather easy to do so. Copp argues that there are many examples of valid arguments that do not have the inference licensing property as Schroeder understands it. Consider for example:

Premise: Everything Jerry thinks is true.

Conclusion: If Jerry thinks Olivia is a dago, then Olivia is a dago.

This argument is valid, however one could both appreciate its validity and be in a position to sincerely assert the premise without at the same time being in a position to sincerely assert the conclusion (and without being irrational). This is because, as will be discussed in more detail later, a fully sincere assertion of the conclusion would require a speaker to feel contempt for Italians. Instead, the property that we should think valid arguments have is the *belief* inference licensing property. That is, valid arguments are such that believing their premises commits one to believing the conclusion—provided the argument's validity is knowable *a priori*. This is a much more plausible thesis, and a hybrid theorist like Copp can endorse it.

Copp notes, however, that "in offering an argument for a moral conclusion, we often want to achieve more than simply to create rational pressure on people to *believe* the conclusion."[54] Often we want to *motivate* them to act in accordance with the conclusion. According to Copp this motivational pressure is the pressure to be in a position to felicitously assert something that you believe. The validity of the argument rationally pressures a believer in the premises to accept the conclusion, and the normativity of the conclusion pressures an assenter to the conclusion to form the appropriate motivation.

[54] Copp (2014), p. 67. Emphasis added.

This brings us to Schroeder's second objection, one that takes aim at the hybrid theorist's understanding of the connection between moral thought and motivation. According to Schroeder, "'It is precisely this idea'—that new moral conclusions have the ability to [independently] motivate—'that is at the heart of the theoretical grounds for the sort of internalism that motivates hybrid theories in the first place.'"[55] In order to establish an internal connection between moral thought and motivation, Schroder thinks that hybrid theories must endorse the Big Hypothesis. If they cannot do this, then Schroeder thinks that the primary justification for pursuing the hybrid strategy in the first place will be undermined. The Big Hypothesis states that if a sentence expresses some group of mental states, then in order for someone to count as believing that sentence, they must be in those mental states.[56]

In the cases that interest us, the Big Hypothesis amounts to the claim that someone can count as *believing* a moral claim only if they are in both the cognitive and motivational states that would be expressed by a sincere assertion of that claim. This would, in effect, entail orthodox motivational judgment internalism: moral belief necessarily implies moral motivation. As noted above, such a view is incompatible with Psychological Humeanism, and thus is unattractive to most hybrid theorists at the very outset. Thus Schroeder takes

---

[55] Copp (2014), p. 69.

[56] This way of speaking betrays a sympathy for what is called "ideationist semantics." Ideationist semantics analyze the meanings of sentences in terms of the mental states of speaker that felicitously use them. I personally follow philosophers of language like Mitch Green and Dorit Bar-On in thinking that it is at best highly misleading to say that *sentences* express mental states. Fortunately, nothing is lost by reinterpreting Schroeder and Copp here to be considering the expression of mental states by speakers who *assert* those sentences. We will return to the issue of ideationist semantics at the end of the next chapter.

himself to be presenting a challenge to the hybrid theorist: they must either give up on one (internalism) or the other (Humeanism) of two important motivations for their view.

Unsurprisingly, Copp doesn't think that hybrid theorists need to accept the Big Hypothesis, or even orthodox internalism, in order to give a satisfying explanation of the connection between moral thought and motivation. According to his theory, Realist-Expressivism, the explanation of Practicality comes from the fact that norms of thought inherit the norms of speech. Here is Copp's example:

> On my version of realist-expressivism, it would be infelicitous for a person to think that torture is wrong—to think of torture as wrong, to have the episodic occurrent thought that torture is wrong—if she lacked a policy of avoiding wrongdoing. In this case she would have a thought it would be infelicitous to express in the straightforward way by asserting "Torture is wrong." So if there are assertability norms governing moral uses of "wrong" such that a speaker who calls something wrong (s)implicates that she has a policy of avoiding wrongdoing, then there would be an infelicity in having the episodic occurrent thought that, say, torture is wrong if one lacked a policy of avoiding wrongdoing.[57]

Copp's point here is suggestive, but like Finlay's comments on "speaking to ourselves" above it obscures the complexity and importance of the work left to be done. For one thing, we need to know exactly what "having an episodic occurrent thought" amounts to—remember, it cannot mean just "having an occurrent belief." But perhaps the more difficult issue is to explain how it is possible for "infelicity" to occur at the *intra*-personal level. These are questions that my positive theory will approach head-on in chapter 4, so we will set them aside for the moment.

[57] Copp (2014), pp. 57–58.

Section 4: Objections to Copp's View

In this section I will focus on two central problems that arise for any hybrid metanormative theory that, like Copp's, attempts to explain a normative judge's expression of motivation by appealing to the meanings of normative terms. The first has to do with the differences in motivation expressed across normative domains. The second has to do with Copp's handling of the disanalogies between normative terms, conventional implicatures, slurs, and pejoratives.

For starters, suppose we are watching *The Sopranos* and observe one of Tony Soprano's mafia henchmen make a terrible blunder. Let's imagine that this underling has a moment of weakness and allows an opportunity to brutally murder a rival slip through his fingers. Caught up in the narrative, I ineffectually call out to the screen "How could you do that? You should have killed the guy!" Here is a response that is not open to you, unless it is meant in jest: "What an awful thing to say! Of course he shouldn't have killed the guy—murder is wrong!" The reason why this criticism would be misguided is that it should be clear from the context that I haven't made a moral claim. Rather, I have used the thin normative term 'should' in a different domain.

This illustrates the general fact that a wide variety of thin normative terms (should/shouldn't, ought/oughtn't, right/wrong, good/bad) may intelligibly be put to use in a wide variety of different normative domains. In the example above, my claim is probably best understood as a sort of prudential claim about what would have advanced the character's interests. If you tell me that my pink jacket is a *bad* one to wear with my orange pants, or that Petit Verdot is the *wrong* sort of wine to drink with Szechuan cuisine, you are probably making aesthetic claims. At the same time, these thin normative terms may

be used in different contexts in the course of advancing other sorts of normative claims like moral claims (e.g. "it is *bad* to cause unnecessary harm"), or epistemic claims (e.g. "it is *wrong* to believe something on insufficient evidence"). When we use the term in a different normative domain, somewhat different motivational or affective states are being espoused. What explains this difference?

Remember Copp's condition (*b*) of conventional simplicature, "the fact that the speaker thereby communicates the proposition *p* is determined by the (or a) conventional meaning of some particular linguistic device in the sentence." This means that any differences in the proposition conventionally simplicated, from one normative domain to another, must be accounted for in terms of the conventional meaning of the thin normative term being used. But how can Copp appeal to a difference in meaning when it looks like the very same sentence is being used in each of the different domains? For example, the sentence 'he shouldn't have done that' may be used to express moral disapproval just as easily as disapproval with respect to etiquette, rationality, or aesthetics.

Given his commitment to (b) I can think of three possible responses to make on Copp's behalf. The first is to bite the bullet and insist that for each individual thin normative term the very same cluster of practical states is expressed, no matter the normative domain in which it is used. The feeling of disapproval I express towards the mafia henchman for failing to advance his interests, is the very same feeling of disapproval that you would express were he to have killed his rival (in both cases we may express this by saying "he *shouldn't* have done that"), despite the fact that one is a prudential claim and the other is a moral one. Or take another example: the very same feeling of approval is expressed when I say that the tawny port is the *right* wine to go with your cheesecake, as when I say that

sacrificing your expensive hobby to set aside money for your child's education is the *right* life decision. I take this proposal to be obviously unappealing, so let's set it aside.

The second option is somewhat more plausible, though it involves taking on some pretty substantial commitments. On this view the usage rules governing the assertability of thin normative terms would be context dependent. That is, one of the rules governing the use of 'should' might be that when you are in a context in which it is clear that you are speaking of prudential norms, you may only felicitously predicate it of an action which you would be—ceteris paribus—somewhat motivated to perform from concern for you own interest. At least, the rule would have to look *something* like that; I must confess I am uncertain about the details. What *is* certain is that pursuing this strategy would involve postulating quite a few more usage rules governing thin normative terms than we otherwise might have thought there to be. For each normative term, in addition to specifying what it contributes to the truth conditions of the sentences in which it occurs, its meaning would also include usage rules specifying its felicitous predication in each of the distinct normative domains in which it might possibly be put to use.

The third possible way of accounting for the phenomenon in question would be to say that for each putative "thin" normative term, there is actually a cluster of as yet unappreciated homonyms. So then, instead of one thin normative predicate 'wrong' that may be used across a wide variety of normative discourses, what we actually have is '$\text{wrong}_{\text{moral}}$,' '$\text{wrong}_{\text{aesthetic}}$,' '$\text{wrong}_{\text{prudential}}$,' '$\text{wrong}_{\text{etiquette}}$,' '$\text{wrong}_{\text{epistemic}}$,' etc.—and so on for all thin predicates and all distinct normative domains. These are each distinct terms with different meanings, though they all sound and look the same.

Such a view could be made more palatable by emphasizing what each cluster of terms has in common. However, this homonym strategy would still have difficulty accounting for uses of thin normative terms that do not belong properly to any particular normative domain, or even any conjunction or disjunction of domains. I'm thinking here, for example, of Bernard Williams's use of 'should' in his discussion of Socrates' question "how should one live?":

> The impersonal Greek phrase translated as *one should* is not only silent about the person whose life is in question. It is also entirely noncommittal, and very fruitfully so, about the kinds of consideration to be applied to the question. "How should I live?" does not mean "what life morally ought I to live?", this is why Socrates' question is a starting point different from those other questions I mentioned, about duty or about a life in which one would be good. It may be the same as a question about the good life, a life worth living, but that notion in itself does not bring in any distinctively moral claims. It may turn out, as Socrates believed and most of us still hope, that a good life is also the life of a good person (*must be* is what Socrates believed; *can be* is what most of us hope). But, if so, that will come out later. *Should* is simply *should* and, in itself, is no different in this very general question from what it is in any casual question, "what should I do now?"[58]

Of course, one can disagree with Williams and Socrates that such uses of thin normative terms are intelligible "because questions such as 'what should I do?', 'what is the best way for me to live?', and so on, are ambiguous."[59] However, such views risk impoverishing our normative speech and thought.

It may be that there are untapped resources available to Copp for dealing with this first issue, so I will give his theory the benefit of the doubt on that account. At the very least, we can see that Copp's theory has not yet tried to address the potential challenges

---

[58] Williams (1985), p. 5.
[59] Williams (1985), p. 5.

that arise from attempting to capture all the various normative domains beyond the moral. Unfortunately, there is a more serious problem facing Realist-Expressivism. This is the problem of embedding. As mentioned already, in his 2014 paper Copp admits that in embedded contexts there are some disanalogies between the way that slurs and conventional implicatures behave on the one hand, and the way that pejoratives and normative terms behave on the other. However, he does not address the significance of this fact or the problem it creates for his Realist-Expressivism.

Let's begin with conventional implicature. Remember, this was the phenomenon to which Copp originally appealed in order to explain Frege expression. Despite the move to conventional simplicature, the expression produced by using these alleged conventional implicature devices (ACIDs) should remain *of a piece* with that produced by other forms of Frege expression. This is because Copp takes conventional implicatures to be a subset of conventional simplicatures—namely, those simplicatures that are part of speaker meaning.

Imagine speakers making these statements:

(5) If all the new students are Canadians and are therefore friendly, then Sally is a Canadian and is therefore friendly.

(6) David believes that Sally is a Canadian and is therefore friendly.

(7) Alicia said that Sally is a Canadian and is therefore friendly.

(8) Consider for the sake of argument that Sally is a Canadian and is therefore friendly.

(9) I'll take a wild guess and say that Sally is a Canadian and is therefore friendly.

In (5) ‘therefore’ is used embedded in a conditional. In (6) it is embedded in a propositional attitude ascription while in (7) it’s embedded in an indirect quotation. In (8) and (9) ‘therefore’ is used in statements that are put forward with non-assertive force—suppositional and speculative respectively. In all five of these examples, the implicature (or simplicature) *scopes out*. That is to say, despite the embedded contexts the speaker still conveys that being friendly follows from being Canadian. At least, that is what is predicted by the traditional view—perhaps some of these examples resist embedding more than others.

It will suffice to say that ACIDs like ‘therefore’ are *fairly good* at scoping out. As it happens, racial and ethnic slurs are even better. Imagine someone saying any of the following:

(10) If all the new students are dagos, then Olivia is a dago.

(11) Jerry believes that Olivia is a dago.

(12) Consider for the sake of argument that Olivia is a dago.[60]

If the slur ‘dago’ is an obscure one for the reader, they are invited to substitute a more familiar one. In each of these three examples the speaker expresses an offensive and harmful attitude, despite the fact that the slur is only used in an embedded context. In contrast, suppose a speaker makes one of these claims:

(13) Eating meat is wrong.

[60] I take it that sort of embedding taking place in propositional attitude ascriptions and indirect quotations is relevantly similar, as is that involved in various kinds of non-assertive force. So from now on three examples should be sufficient.

(14) If eating meat is wrong, then eating this hamburger is wrong.

(15) David believes that eating meat is wrong.

(16) Consider for the sake of argument that eating meat is wrong.

Any expression of disapproval towards eating meat that might be expressed by a speaker who asserts (13) disappears the moment the moral claim is embedded as in (14)–(16).

What about thick normative terms, do they behave differently than thin ones? Consider the following utterances:

(17) If all the new students are cowards, then Jack is a coward.

(18) Sally believes that Jack is a coward.

(19) Consider for the sake of argument that Jack is a coward.

At first it is hard to see how the speaker of (17)–(19) thereby conveys any feeling of disapproval. However, the case is a bit more complicated than it was for thin terms. This is because the speaker who uses a thick term like 'coward' (even if she does so in an embedded context) thereby conveys that she thinks the word *may sometimes be applied correctly*. This means that the speaker conveys that she thinks there *is* a character trait one *could* have, such that one displays a morally objectionable tendency to excessive fearfulness. Thus disapproval of cowardice is expressed at a general level by the speaker of (17)–(19), even if the speaker expresses no attitudes towards Jack in particular. So thick normative terms don't behave *exactly* the same way thin ones do. Does this spell hope for Copp's analogy between normative terms and slurs?

To evaluate the situation, let's look at the toughest test case—one in which it seems most plausible that a thick normative term scopes out. Take a thick normative term that most of us would feel uncomfortable using:

(20) If all the new students are unchaste, then Ava is unchaste.

(21) Simon believes that Martha is unchaste.

(22) Consider for the sake of argument that Riley is unchaste.

Under normal circumstances, the speaker in (20)–(22) seems to convey disapproval. They do not convey disapproval towards the particular individuals to which unchastity is being predicated in the embedded context, however they *do* convey disapproval towards certain general patterns of sexual activity (namely, that they may be evidence of a vicious character trait). Is this not perfectly analogous to what is happening in the slur cases? After all, in (10)–(12) the speaker does not convey an offensive attitude towards the particular subjects of whom the slur is predicated in an embedded context. Rather, the speaker in (10)–(12) expresses an offensive attitude toward Italians in general—that it may be sometimes appropriate to apply the term 'dago' to them.

In fact, I do not think that the cases are perfectly analogous. This is because a speaker may cancel the implicature generated by the embedded use of a thick normative term by making it explicit that they do not think such a term could ever be used appropriately. This is not true of slurs. Consider:

(23) If all the new students are unchaste, then Ava is unchaste. Just to be clear though, I don't think anyone is unchaste because I do not believe there is any such vice.

The speaker in (23) cancels the implicature that they disapprove of a particular pattern of sexual behavior as evidence of a vice, without any residual absurdity. Of course, we may still need some background context to know what the point was of their asserting the conditional. But that is no more problematic than a statement like

(24) If all the animals at the zoo are unicorns, then Patches is a unicorn. Just to be clear though, I don't think there are any unicorns because I do not believe there is any such animal.

We might need some background context to understand *why* the speaker of (24) asserted the conditional, but even without additional explanation their cancellation involves no absurdity. The audience is left with no lingering sense that language has been used incorrectly. The same is true of (23). Similarly, consider what it would sound like to cancel (17):

(25) If all the new students are cowards, then Jack is a coward. Of course I don't think anyone is a coward because I do not believe there is such a vice.

Again, though we might be *interested* to hear more from the speaker of (25), the audience will not be linguistically unsettled by what is said even if no further explanation is forthcoming.

Let's extract a more general lesson from these observations. When a speaker performs a predication in an embedded context, often they thereby indicate that they think the predicate may sometimes appropriately be applied in *unembedded* contexts. Of course, this is not always the case. For example, there is a general presumption that speakers do not believe that there are any unicorns. In many modern contexts there may even be a presumption that the speaker does not believe there is such a vice as unchastity. In the case

of many other thick normative terms like ‘cowardice’ there is probably a general presumption that speakers *do* believe there is such a virtue or vice. Thus, all else being equal, speakers who predicate thick terms like ‘cowardice’ of subjects in embedded contexts will thereby convey that they believe there are such virtues and vices. When a speaker expresses *that a predicate may sometimes be applied correctly* simply by using it in an embedded context, I believe this is a good example of a generalized conversational implicature. Particular features of a conversational context are not required to generate the expression (though they could preempt it), and it is easily cancellable without absurdity.

This is decidedly *not* what is going on in the case of slurs. Consider for example:

(26) If all the new students are dagos, then Olivia is a dago. Just to be clear though, I don’t think anyone is a dago because I do not think dago is a term that is ever appropriately applied.

A speaker who asserts (25) does not accomplish a free and clear cancellation anywhere near the magnitude of intelligibility present in (23)–(25). This is because slurs are terms that are colored with contempt in such a way that it is extremely difficult for a speaker to distance herself from them. For this reason we often hesitate to utter slurs, even when it is clear we are only *mentioning* them (rather than *using* them). If we must do so, it is usually only after a careful preamble or disclaimer. Even in the most sanitized and academic of contexts, a placeholder will often be used instead (e.g. “the ‘N’ word,” “the ‘C’ word,” etc.).

To make the difference between the normative cases and the slur cases vivid consider what happens when we consider the canceled version of (21):

(27) Simon beliefs Martha is unchaste. Just to be clear though, I don't think anyone is unchaste because I do not think there is such a vice as unchastity.

In (27), the implication that 'unchastity' may sometimes be predicated appropriately is pretty well cancelled. At least, the speaker of (27) is able to successfully distance herself from it. Compare this with the embedded use of a gender slur:

(28) If all the new students are sluts, then Ava is a slut. Just to be clear though, I don't think anyone is a slut because I do not think slut is a term is ever appropriately applied.

To my ears, the speaker of (28) digs himself into a deeper and deeper hole with every utterance of the slur in question. This provides a clear contrast with (27).

In fact, this very feature of slurs was taken advantage of by Donald Trump during his 2016 primary campaign for the republican presidential nomination. At a Trump rally in Manchester, New Hampshire on February 8th 2016, while Donald Trump was talking about Senator Ted Cruz's stance on torture, a supporter near the front of the crowd yelled out "Pussy!" clearly referring to Cruz. Here is the response that Trump made during his speech as the situation unfolded:

> She just said a terrible thing. You know what she said? Shout it out 'cause I don't want to. Ok. You're not allowed to say, and I never expect to hear that from you again. She said—I never expect to hear that from you again! She said he's a "pussy!" Terrible. Terrible. That's terrible.[61]

Trump here is able to clearly express the very same attitudes as the supporter from the crowd, by embedding a slur in a quotation *and* at the same time explicitly denying that he

[61] Spoken Werd (2016).

himself would say such a thing. He mentions the slur rather than using it, fully aware that the charged nature of the word is on its own enough produce the effect he desires. His protestations that using the word would be "terrible" serve as a perfect illustration of the fact that cancelling expression in such cases bound to be unsuccessful. Trump knows and exploits this fact.

Before we move on, let's examine one final difficult case. Consider light pejoratives. For example:

(29) Jerry is a jerk.
(30) If all the new students are jerks, then Jerry is a jerk.
(31) Alicia believes that Jerry is a jerk.
(32) Consider for the sake of argument that Jerry is a jerk.

Once again, any expression of contempt towards Jerry (29) evaporates when the claim is embedded (30)–(32). One can easily imagine a speaker of (30)–(33) even following their utterance up with a successful denial that they themselves would use such a word as 'jerk' at all. It is significant that there is a division here between the embedded behavior of ACIDs and slurs on the one hand, and the embedded behavior of normative terms and light pejoratives on the other.[62] Thick normative terms, thin normative terms, and light pejoratives simply do not have the same coloring attached them that slurs and ACIDs do. This indicates that the sharp division in modes of expression is still most naturally drawn with the ACIDs and slurs on one side and the thick and thin normative terms on the other.

I propose that this is not, as Copp would say, merely a matter of different terms having somewhat different usage conditions. Rather we are running up against two

[62] For more on scoping out in general and this division in particular, see Schroeder (2014) and Hay (2013).

different mechanisms of expression. In the case of conventional implicatures and slurs, Copp's view may be plausible. Coloring in general, and hence Frege expression, scopes out in embedded contexts and the resulting expression resists cancellability. However, whatever the mechanism of expression in the normative case, it is one that does not survive embedding (and even in cases where there is a hint of scoping out, such expression is easily cancellable).

A natural question to ask at this point is why light pejoratives like 'jerk' behave in a way that is similar to the normative terms. I believe the answer is that they just *are* also normative terms. In the case of 'jerk' what we have is a thick normative term that is most naturally at home in discourses about ethics or perhaps etiquette. But there are many other examples as well. Think of the aesthetic term 'philistine' or the prudential term 'fool.' These are all negative cases, but I think there are many positive ones as well (e.g. 'cool,' 'savvy,' 'nice,' etc.). There are many normative terms that we effortlessly apply in a variety of normative domains without a second thought, both terms of criticism and terms of praise. What they all have in common with normative statements in general is that they involve a mechanism of expression that does not survive embedding.

My suggestion is that slurs, pejoratives, approbatives, laudatives, epithets, and the like will all fall into one or the other of two categories. That is, they will either scope out in embedded contexts or they won't. If they do scope out, and the resulting expression is not easily cancellable, then that is strong evidence that they are colored terms and good candidates for Frege expression. If they do not scope out (or if any lingering expression from the embedded context is easily cancellable), then they are probably mere normative terms from one domain or another. As I have been hinting, the related phenomenon worth

investigating here involves the connection between scoping out and detachability. ACIDs and slurs all have corresponding neutral terms that may be used in their place (e.g. 'and' for 'but', 'Italian' for 'Dago', etc.), while normative terms and light pejoratives do not. It could be that part of what explains the scoping out behavior of the former is the fact that a speaker is explicitly choosing to use the colored term rather than the neutral counterpart.[63] This may be good news for the future prospects of Copp's analysis of a restricted class of slurs.[64] However, it is bad news for his analysis of normative terms.

Before moving on, I would like to apologize to the reader for my uses of slurs in this section to illustrate my points. I do not do so lightly. One part of me would have preferred to preface this section of the chapter with a disclaimer, while the part that won out believes that doing so would have robbed the examples of their rhetorical power. To borrow a phrase from the previous chapter, I find my own normative commitments at loggerheads with one another. As is often the case when theorizing about the behavior of offensive language, the value of constructing and arguing for an adequate account competes against the potential disvalue of harming one's audience and even oneself. I am especially sensitive to this fact given that, if I what I have said is right, the expressive content of such language can bleed through even the most careful attempts to contextualize and embed it.

---

[63] Interestingly, if this hypothesis is plausible, then it would seem to imply some things about thick normative terms. Namely, if the thesis that thick normative terms are cleanly separable into descriptive and normative contents is true, then thick normative terms would be detachable. That is, their neutral counterparts are already present in the natural language. Thus if we intuitively think that thick normative terms *do not* scope out in embedded contexts, and hence are *not* detachable, then this would cast doubt on the "clean separability" thesis.

[64] For a closer look at this view, see Sennet & Copp (2014).

Section 5: Taking Stock

There is much in Copp's view that I can agree with, and still more to learn from. He is right that speakers who put forward sentences containing normative terms with the right sort of speech act force are able thereby to express both cognitive and motivational states. This expression is noncancellable, detachable, and does not contribute to the truth conditions of what the speaker has said. We have learned more about what it takes for a hybrid theory to do justice to the spirit of motivational judgment internalism without contradicting Psychological Humeanism, and that it is possible to make sense of the way that valid arguments containing normative terms put pressure on a reasoner to infer true conclusions from true premises, while at the same time explaining the fact that we use these arguments to motivate each other was well.

We are also in a position to diagnose what led Copp astray. Any view that tries to make sense of the mechanism of expression involved as grounded in the *meanings* of normative terms will be inadequate. This includes theories that chalk the expression up to semantic content, usage conditions, conventional implicatures, conventional simplicatures, Fregean coloring, and the like. Such forms of expression have a habit of scoping out of embedded contexts in a way that the expression of motivation (or feeling) generated by the use of normative terms does not. In the next chapter I will begin to lay out the details of my positive view of normative language, one that avoids the vices of both Finlay's and Copp's views while combining their virtues.

CHAPTER 3

A TRULY HYBRID THEORY OF NORMATIVE SPEECH

Part I: Introduction

In this chapter I explain the details of my positive view of normative speech. My goal is to solve the puzzle identified at the end of the last chapter: How can we maintain that speakers who make normative claims express motivational or affective attitudes in a way that is noncancellable, without running afoul of the broad problem of embedding? Or, put another way: How may we maintain that speakers who make normative claims only express motivational or affective attitudes when in particular circumstances, without committing ourselves to the notion that they may also felicitously cancel this expression at will? Any so-called "hybrid" view that is forced to compromise—to give up on noncancellability to solve the broader problem of embedding or vice versa—hasn't truly "had it both ways."

In the next part of this chapter, I devote some space to the interplay between speech acts, norms, and varieties of expression. I describe how constitutive norms individuate one speech act from another, allow speech acts to play important functional roles within

different kinds of conversations, and give speech acts their expressive power. Along the way I appeal to important insights from Austin, Sellars, Grice, Dorit Bar-On, and Mitch Green. The picture that emerges is one of speech acts as integral parts of our social practices that enable us to stick our necks out to one another, and thereby to show each other our inner states.

In Part III I show that, given this better understanding of speech acts, it is natural to claim that when we are speaking from within different normative domains, we are really making use of similar but distinct speech act types. According to my view normative speech acts are united by their requirements that speakers using them are in both cognitive and moving states of mind. Different normative speech act types are differentiated from one another by variations in the precise contours of those states of mind—hence the similarities and differences in what speakers are able to express to one another by making use of these speech acts and the functional roles these acts are able to play from within their respective normative domains of discourse. Ultimately, the differences between one normative speech act type and another are explained by the differences in aims of the various domains of normative practice in which they arose. Along the way I argue for a more general method that can allow us to determine when we should posit new speech act types.

I then apply my hybrid view to particular cases in the ethical and aesthetic domains. First, I show how a hybrid speech act view of ethical discourse is able to explain a longstanding metaethical puzzle. In rough terms this puzzle is captured by Michael Smith's moral problem: the apparent inconsistency between metaethical cognitivism, moral judgment internalism, and Psychological Humeanism. Next, I show how a hybrid speech

act view of aesthetic discourse is able to explain the puzzle of aesthetic judgment. This puzzle consists in the apparent inconsistency of three independently plausible claims about the norms of assertion, aesthetic testimony, and proper aesthetic conversations.

The chapter concludes in Part IV with a brief discussion of the relationship between my theory and two close competitors (Dorit Bar-On and Dan Boisvert), and a summary of where we are so far. The positive evidence for my own view consists in its coherently paving a way between Finlay and Copp—and thus truly "having it both ways"—together with its promising ability to solve outstanding puzzles in multiple normative domains. However, as it stands this victory should still ring hollow. This is because the phenomenon of true interest for many engaged in these debates (including myself!) is not normative speech at all but normative thought. I briefly describe the explanatory burden that remains to be discharged, in order to set the stage for the next chapter.

Part II: Speech Acts and Self Expression

Section 1: Austin on Speech Acts

In his 1955 William James lectures at Harvard, published as *How to Do Things with Words* in 1962, J. L. Austin describes his target phenomena as being "widespread and obvious" and as that which "cannot fail to have been already noticed, at least here and there by others."[65] From the perspective of the present, Austin's observations have been so influential on philosophy of language, pragmatics, and linguistics, that reading his book

[65] Austin (1962), p. 1.

now feels more like an exercise in making one's own assumptions explicit than in revisiting a theory of the past. Nevertheless, this is still a worthy exercise—and it is where we will begin.

> We shall take, then, for our first examples, some utterances which can fall into no hitherto recognized *grammatical* category save that of 'statement', which are not nonsense, and which contain none of those verbal danger-signals which philosophers have by now detected or think they have detected (curious words like 'good' or 'all', suspect auxiliaries like 'ought' or 'can', and dubious constructions like the hypothetical): all will have, as it happens, humdrum verbs in the first person singular present indicative active. Utterances can be found, satisfying these conditions, yet such that
>
> A. they do not 'describe' or 'report' or constate anything at all, are not 'true or false'; and
> B. the uttering of the sentence is, or is a part of, the doing of an action, which again would not *normally* be described as, or 'just', saying something.
>
> This is far from being as paradoxical as it may sound or as I have meanly been trying to make it sound: indeed, the examples now to be given will be disappointing.
>
> Examples:
> (E. a) 'I do (sc. Take this woman to be my lawful wedded wife)'—as uttered in the course of the marriage ceremony.
> (E. b) 'I name this ship the *Queen Elizabeth*'—as uttered when smashing the bottle against the stem.
> (E. c) 'I give and bequeath my watch to my brother'—as occurring in a will.
> (E. d) 'I bet you sixpence it will rain tomorrow.'
>
> In these examples it seems clear that to utter the sentence (in, of course, the appropriate circumstances) is not to *describe* my doing of what I should be said in so uttering to be doing or to state that I am doing it: it is to do it. None of the utterances cited is either true or false: I assert this as obvious and do not argue it. It needs argument no more than that 'damn' is not true or false: it may be that the utterance 'serves to inform you'—but that is quite different. To name the ship *is* to say (in the appropriate circumstances) the

> words 'I name, &c.'. When I say, before the registrar or altar, &c., 'I do', I am not reporting on a marriage: I am indulging in it.[66]

We will return in Part III to those "curious words" that Austin sets aside. For now we are primarily concerned with Austin's distinction between the different components of a given speech act. Specifically, between the sentence uttered and the act performed by the speaker *by so uttering*. Importantly for us, Austin clearly highlights the fact that the grammatical structure of the sentence underdetermines which speech act it is being used by the speaker to perform. We may go even further by pointing out that the semantic content of the sentence *also* underdetermines which act it is being used to perform. Though Austin's emphasis in this passage is on "performative" speech acts, we may take note that his distinction holds equally true of other speech acts as well.[67]

To see this, just take Austin's last example:

(E. d) 'I bet you sixpence it will rain tomorrow'

And compare the prediction:

(E. e) 'I predict that it will rain tomorrow'

Of course, given the right context, I could use the sentence I uttered in example *d* to make the very same prediction I did in example *e*. That is, I could perform the speech act of *prediction*, without explicitly saying that I am doing so. Suppose this is what I am doing when I say:

---

[66] Austin (1962), pp. 4–6.
[67] Austin (1962), p. 20.

(E. f) 'I bet you sixpence it will rain tomorrow'

Given that "sixpence" is an unusual sum of money to be exchanging these days, were I to utter *f* you would probably assume that I was only saying what I did in order to offer a quirky sounding prediction rather than asking you to make a bet with me. Notice that the difference between *d* and *f* is not a difference in the *vehicle* (sentence) used to perform the speech act, but a difference in the *type* of speech act performed.[68] In recognizing this, we admit to ourselves that the description of the background circumstances (the conversation in which the speaker makes a prediction vs. the conversation in which the speaker places a bet) is an ineliminable part of the example.

If a vehicle may remain identical across different speech acts—i.e. the same sentence, with the same words arranged in the same way—what is it that differentiates one speech act from another? To say that they are different "acts," or that they are utterances of the same sentence with different "forces" is merely to rephrase the *explanandum*, not to offer an *explanans*. So far we have discussed the influence of differences in context, but this is still much to vague. Fortunately, Austin helps us out here as well:

> Let us first at least concentrate attention on the little matter already mentioned in passing—this matter of 'the appropriate circumstances'. To bet is not, as I pointed out in passing, merely to utter the words 'I bet, &c.': someone might do that all right, and yet we might still not agree that he had in fact, or at least entirely, succeeded in betting. To satisfy ourselves of this, we have only, for example, to announce our bet after the race is over. Besides the uttering of the words of the so-called performative, a good many other things have as a general rule to be right and to go right if we are to be said to have happily brought off our action. What these are we may hope to discover by looking at and classifying types of case in which something *goes wrong* and the act—marrying, betting, bequeathing, christening, or

[68] 'Vehicle' is Dorit Bar-On's term. See her (2004) chapter 6, especially pp. 217–225. More on vehicles below.

> what not—is therefore at least to some extent a failure: the utterance is then, we may say, not indeed false but in general *unhappy*. And for this reason we call the doctrine of *the things that can be and go wrong* on the occasion of such utterances, the doctrine of the *Infelicities*.[69]

"Infelicities" are meant to cover both what is going wrong when a performance *fails entirely to count as a speech act of the type intended*, and what is going wrong when a performance *succeeds in being a speech act of the type intended, but is unhappy in some other way.* For a clear example of the former, one need only (as Austin says) attempt to place a bet after the race is over—merely saying the words is not enough. Austin calls these failures *MISFIRES*. For a clear example of the latter, consider the case of lying. An act of lying is constituted by a successful performance of an assertion that a speaker believes to be false and yet performs anyway, usually with the intent to deceive. Such a token of the type *assertion* is a paradigm example of infelicity. Austin calls these "unhappy" performances *DISREPSECTS.*[70] Are the possibilities for *DISRESPECT* exhausted by failures of sincerity? They are not.

For one thing, Austin includes in the category of *DISRESPECTS* any failure to follow through on a speech act that involves committing oneself to future conduct. For example, if one makes a promise in good faith, the subsequent failure to follow through on one's word would be considered an instance of *DISRESPECT*, even though the promise was made in all sincerity. Austin offers us the following useful list of rules with an accompanying chart:

---

[69] Austin (1962), pp. 13–14.

[70] In the main text Austin uses the term *ABUSES* to refer to this subset of infelicities. I prefer *DISRESPECTS*, also a term that Austin used for this category, as it is less charged. See Austin (1962), p. 18 footnote.

(A. 1) There must exist an accepted conventional procedure having a certain conventional effect, that procedure to include the uttering of certain words by certain persons in certain circumstances, and further,

(A. 2) the particular persons and circumstances in a given case must be appropriate for the invocation of the particular procedure invoked.

(B. 1) The procedure must be executed by all participants both correctly and

(B. 2) completely.

(*Γ*. 1) Where, as often, the procedure is designed for use by persons having certain thoughts or feelings, or for the inauguration of certain consequential conduct on the part of any participant, then a person participating in and so invoking the procedure must in fact have those thoughts or feelings, and the participants must intend so to conduct themselves, and further

(*Γ*. 2) must actually so conduct themselves subsequently.[71]

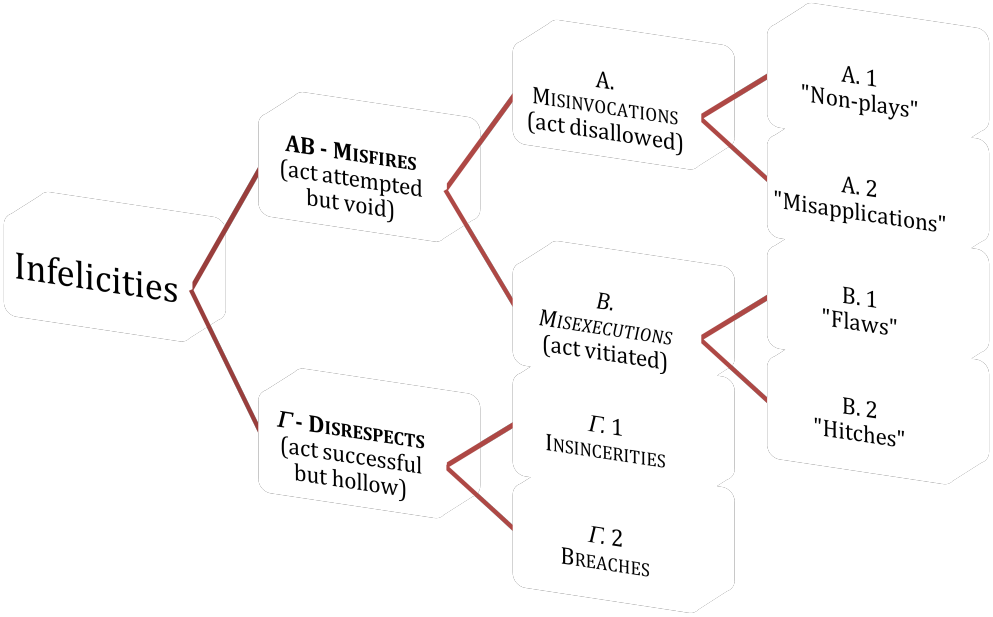

Figure 1: Austin's Infelicity Scheme[72]

In what follows I will have little to say about *BREACHES* (violations of *Γ*. 2), and a great deal to say about *INSINCERITIES* (violations of *Γ*. 1). However, I should like to clear up a

[71] Austin (1962), pp. 14–15.

[72] From Austin (1962), p. 18 and completed using his footnote 1.

common misconception about the class of insincerities. It has been suggested by some that breaking the rules that fall into this category is best thought of as involving intentional deception.[73]

As discussed above, the simplest and clearest case of insincerity—that of lying—certainly does typically involve intentional deception. However, violating the insincerity rule (*Γ*. 1) need not always involve the intention to deceive. For example, given Austin's categorization of infelicities it is quite clear that we should take the requirement that asserters believe what they are asserting as a condition of *SINCERITY* so construed. That is to say, "the procedure" (asserting) "is designed for use by persons having certain thoughts" (beliefs). It may be that there are more conditions required to bring off the performance of a truly "happy" assertion (such as justification, knowledge, or the reasonable belief that one knows). At the very least, though, assertions should certainly be counted *unhappy* when performed by speakers who do not believe them. Notice that this condition may sometimes be failed by speakers without any intention to deceive whatsoever, but rather by failures of self-knowledge or carelessness.[74] Thus '*INSINCERITY*' in this context is something of a technical term.

So then, we have at last some resources for distinguishing speech act types from one another. When we instantiate Austin's list of *infelicities* to a particular speech act type,

---

[73] For example, see Copp (2009), p. 174. More on this issue in Part III below.

[74] Some might ask why, given these considerations, I have not reverted to Austin's other name for the category *DISSIMULATIONS*. The uncharacteristic inelegance and obscurity of this word is reason enough. Besides, I believe that the common meaning of "insincerity" is sufficiently broad to stand for a wider range of failures than just intentional dishonesty.

we get the complete sets of rules conventionally attached to that act.[75] These fully specified sets of rules serve to *individuate* the speech act types—to differentiate them from one another. Though two different types of speech act might overlap somewhat in their infelicities, a complete and exact overlap would spell speech act *type identity*. When we say that rules governing a speech act type are attached by convention, that is really to say that the speech act *is itself a conventional act*: the felicity conditions governing a speech act type are *constitutive of that action type*.

Of course, other modes of evaluation besides constitutive norms may appropriately be brought to bear on a speech act *token* in a particular set of circumstances. Austin puts this point well:

> The first thing to remember is that, since in uttering our performatives we are undoubtedly in a sound enough sense 'performing actions', then, as actions these will be subject to certain whole dimensions of unsatisfactoriness to which all actions are subject but which are distinct—or distinguishable—from what we have chosen to discuss as infelicities. I mean that actions in general (not all) are liable, for example, to be done under duress, or by accident, or owing to this or that variety of mistake, say, or otherwise unintentionally. In many such cases we are certainly unwilling to say of some such act simply that it was done or that he did it. I am not going into the general doctrine here: in many such cases we may even say the act was 'void' (or voidable for duress or undue influence) and so forth. Now I suppose some very general high-level doctrine might embrace both what we have called infelicities *and* these other 'unhappy' features of the doing of actions—in our case actions containing a performative utterance—in a single doctrine: but we are not including this kind of unhappiness—we may just remember, though, that features of this sort can

[75] I should be clear that I am not necessarily committed to a conventionalist theory of promises. I am, however, committing myself to a conventionalist theory of the *speech act* of promising. For example, T. M. Scanlon could be right that the *moral wrongness* of breaking a promise does not depend on an existing social convention, but rather on what we owe to people when we cause them to form certain expectations. In such a case, I would say that it is the conventional rules that constitute the *speech act* of promising that enable it to function as a tool for causing people to form certain expectations. See Scanlon (1998), chapter 7.

> and do constantly obtrude into any particular case we are discussing. Features of this sort would normally come under the heading of 'extenuating circumstances' or of 'factors reducing or abrogating the agent's responsibility', and so on.[76]

Like Austin, I will refrain from getting too far into the general theory here. However, I will add two important observations. The first is to take note of the *reason* for cordoning off these forms of "unsatisfactoriness" from infelicities in particular. The reason is that infelicities are generated by rules that are *constitutive* of the speech act *types*, whereas the forms of unsatisfactoriness identified here are applicable to particular act *tokens* given specific contexts.

My second observation is that, like all action tokens, speech acts may be subject to a whole host of other norms given the particular situation in which they are performed. These are also non-constitutive norms. They include the ethical, aesthetic, prudential, epistemic, rational, etc. considerations in favor of or against performing a particular speech act in a particular set of circumstances, *independently of* its constitutive norms. For example, in particular situations there may be all sorts of grounds for criticizing or praising someone's act of promising ("Your promise was so polite and brave, though perhaps imprudent."). These would invoke non-constitutive norms. In contrast, in order to understand someone as *having made* a promise *in the first place* you must understand their act to be criticizable if it is found out that they had no intention whatsoever to do what they have promised ("You have promised in bad faith!"). This is why the requirement that a speaker intend to follow through is a *constitutive* norm of promising.[77]

---

[76] Austin (1962), p. 21.

[77] There is a perfect example of this in the first season of HBO's series *Game of Thrones* (episode 6). Sean Bean's character, while officially sitting proxy for the king,

Section 2: Bar-On and Sellars on Expression

So far I have introduced a picture of speech acts as acts performable by a speaker merely by *meaning* to do so in the right circumstances. Differences in speech act type are individuated by the constitutive felicity rules that govern what counts as a fully successful performance of the respective type. The next step is to explain how it is that performing certain speech acts enables us to *express* certain mental states (rather than merely describe ourselves as being in them, or give evidence that we are in them). I will accomplish this explanation over the next two sections.

In his 1969 paper "Language as Thought and as Communication," Sellars divides his exploration of the various relationships between thought and language into two sections. The first addresses the extent to which linguistic behavior *is itself a form of* conceptual activity. The second section covers the different ways in which language may be described as "expressing" thoughts. It is this latter issue that is most relevant to my present discussion. Specifically, I will make use of Sellars' distinction between three different senses of "expression:"

> The term 'express' in contexts pertaining to thought has two radically different senses. The difference can be brought out by relating these senses to two different contexts, namely,
>
> (1) Jones expressed his thought (belief) that-p by saying …
> (2) *Jones'* utterance of 'p' expressed his thought that-p

---

gives an order in the king's name. Afterwards, Aiden Gillen's character subtly warns him "A bold move, my lord, and admirable. But is it wise…?" Regardless of how bold, admirable, or unwise the order was it still satisfies all conditions necessary for being a felicitous action of its type. In other words, the constitutive norms of the speech act of "ordering in the name of the king" have been followed, even if *all things considered* the order was inadvisable.

> I shall call the former the 'action' sense of express, and the latter, for want of a better term, the 'causal' sense. Both, as we shall see, are to be distinguished from a third sense illustrated by the context
>
> Jones' utterance of 'p' expressed *the* thought that-p
>
> where the phrase '*the* thought that-p' stands for an abstract entity, a thought in Frege's sense (i.e. in one sense of this term, a 'proposition'). I shall call this the logical (or semantical) sense of 'express.'[78]

Dorit Bar-On puts Sellars' definitions this way:

> **Expression$_1$** (the *action* sense): a *person* expresses a state of hers by intentionally doing something.
>
> **Expression$_2$** (the *causal* sense): an *utterance* or piece of behavior expresses an underlying state by being the culmination of a causal process beginning with that state.
>
> **Expression$_3$** (the *semantic* sense): e.g., a *sentence* expresses an abstract proposition, thought, or judgment by being a (conventional) representation of it.[79]

In the action sense, 'expression' picks out a relation between an agent performing an intentional action and one of her states—usually a mental state. In the causal sense, 'expression' picks out a relation between two events: the activation of a standing disposition being a cause and the event of uttering being an effect.

Notice that these two relations often come together, but can be separated. For example, in most cases of ordinary, fully felicitous assertions a speaker intentionally expresses$_1$ her belief by uttering an indicative sentence—which utterance is at least in part caused by (expressive$_2$ of) the speaker's having the belief—and meaning that utterance as

---

[78] Sellars (1969), pp. 520–521.
[79] From Bar-On (2004), p. 216.

an assertion. However, it is possible to exhibit one form of expression without the other. As Bar-On points out, "an involuntary twitch in my eye may express my discomfort in the causal sense, without my having expressed discomfort by it."[80] One may even be involuntarily caused to utter a sentence as a causal result of having a belief—an episode that will most likely give evidence to those within earshot of your mental state—but without intentionally *meaning* your utterance as an assertion. Freudian slips, if there truly are such things, would be the paradigm cases of such expression$_2$ without expression$_1$. It should be admitted, however, that expression$_1$ is still a very broad category. It includes every occasion in which an intentional action is expressive of a state of mind, even if the actor does not *intend* to express that state of mind *by so acting*. In our discussion of Green in the next section we will zero in on a narrower range of cases of expression$_1$—cases in which the actor intends to engage in an action partly *because* it is designed to be expressive.

Whether there can be expression$_1$, even in the broad sense, without expression$_2$ is an interesting question. It depends on whether one may intentionally perform an action that is expressive of one's state of mind, without one of the causes of that performance being the state of mind in question. This is an empirical question. However, we may note that one most certainly can *purport* to express$_1$—or "represent oneself as expressing$_1$"—without at the same time expressing$_2$. After all, when a speaker tells a lie (provided that they know their own mind) they represent themselves as having a belief that they do not in fact have. It should be noted that expression$_1$ is usually taken to be a success term—one can only express states that one is actually in. For clarity's sake, in this chapter

[80] Bar-On (2004), p. 216.

I will continue to use it as a success term (distinguishing between "expressing$_1$" and "purporting to express$_1$").[81]

The third, semantic sense of expression picks out a relation between a sentence and a proposition. Given this important distinction between the *action* and *semantic* types of expression, we are able to add another chapter to Austin's story: one thing we are often up to when we perform speech acts, is that we are using a sentence that (semantically) expresses$_3$ a proposition to (action) express$_1$ a state of mind. In many cases both the sentence and the state of mind will take the same proposition as their object (as in sincere assertions). But often they will not. When you take me up on my bet by saying "Alright, you're on! I bet you sixpence it *won't* rain tomorrow" the sentence you have uttered expresses$_3$ the proposition *that you have placed a bet in the amount of sixpence to the effect that it will not rain tomorrow*. But the state of mind you have expressed to me by sincerely performing this speech act is one of *intention* that takes as its object the proposition *that you will pay me sixpence if it does rain tomorrow and collect sixpence from me if it does not*.[82]

Alternatively, as observed above, I could utter the same sentence as a vehicle in the performance of a different speech act—say, one of prediction. Suppose we are discussing the weather and I say casually to you "I bet you sixpence it will rain tomorrow" and it is

---

[81] Here I use Green's terminology (see below). In the same spirit, Bar-On also helpfully distinguishes two sorts of "purporting": "pretending to express" and "trying but failing to express." See Bar-On (2004), p. 323–328.

[82] It is an interesting question exactly what state of mind I express to you when I offer the bet initially. Certainly, I do not express the same simple intention—rather, I intend to commit myself to the respective payouts only *conditionally* on you accepting my bet. Betting is a curious, though not unique, sort of speech act in that it depends on the cooperative behavior of more than one individual in order to come off felicitously in the first place. Marrying is another example that comes to mind.

clear to you that I have no intention whatsoever to place a bet, but only to use this turn of phrase to playfully give voice to my prediction. In such a case, my sentence still semantically expresses the proposition *that I have placed a bet with you in the amount of sixpence to the effect that it will rain tomorrow*. However, in such a case I will have expressed to you my belief (of low to moderate credence) in the proposition *that it will rain tomorrow*. The upshot is that while expression$_3$ is determined by the semantic and syntactic properties of the sentence uttered, expression$_1$ is determined by the felicity conditions of the action intentionally performed.

In working towards a neo-expressivist theory of avowals in her book *Speaking My Mind*, Dorit Bar-On picks up where Sellars left off. Here, for example, is what she has to say about action expression$_1$ in cases of assertion:

> Consider first avowals of beliefs. Saying (or thinking) "I believe John is angry with me" is not self-verifying. Still, if it is a case of avowing, the point of issuing the self-ascription seems to be at least in part to give direct voice to my (first-order) belief—that John is angry with me. If so, this could help explain the anomaly of Moore sentences such as "I believe that John is angry with me, but John is not angry with me", of which we spoke earlier. The sentence does not involve an overt contradiction. Its two conjuncts semantically express propositions that are mutually compatible. However, if we consider an utterance of the Moore sentence, we get some conflict. Given the right background conditions, a sincere utterance of a sentence which semantically expresses the proposition that p will typically count as expressing$_1$ (as well as expressing$_2$), one's belief that p. So, when uttering the second part of the Moore sentence ("John is not angry with me"), I will typically be taken to express$_1$ my belief that John is not angry with me. But now suppose that we take the first part of the pronouncement to be an avowal, rather than a third-person theoretical report on my cognitive economy. Then, on my proposed account, we can see it as also expressing$_1$ my belief that John is angry with me. So we get an expressive conflict: in one act I am expressing$_1$ two conflicting beliefs, even though, semantically speaking, the sentence I utter does not express$_3$ a contradiction.[83]

[83] Bar-On (2004), pp. 217–218.

It does seem right to say that this expression occurs, but what exactly does it amount to? Bar-On points out that, in the special case of ascribing to oneself a *thought*, merely uttering a sentence that semantically expresses$_3$ the same proposition as that thought is sufficient for also expressing$_{1+2}$ (in the action and causal senses) that one is entertaining that thought. However, most cases are not like this:

> Merely spelling out of the propositional content or intentional object of one's hope, wish, fear, etc. does not suffice for being in a state of hoping that p, wishing for x, being afraid of y, etc. So, while the point of articulating content when avowing hoping, or wishing, or fearing, may still be to express one's first-order intentional state, the articulation by itself does not guarantee that one will succeed in expressing one's hope that p/wish for x/fear of y etc.; so the avowal is not self-verifying.[84]

How is it that non-self-verifying instances of utterances express a speaker's state of mind? Bar-On's response is to adopt the view she calls "Neo-Expressivism." On this view avowals are "expressive acts in which subjects directly give voice to, by way of sharing, airing, or simply venting, a self-ascribed mental state."[85] Neo-Expressivism offers insights into other linguistic acts of expression beyond avowals as well.[86] The more general application is to cases in which a speaker uses language to intentionally give expression to her mental state, whether or not she does so by explicitly avowing. In such cases, Bar-On distinguishes between the expressive *act* and the *vehicle* of that act. While the *act* may be

[84] Bar-On (2004), p. 221.
[85] Bar-On (2004), p. 227.
[86] For explicit mentions of the link between avowals and other intentional linguistic acts of expression, see Bar-On (2004), p. 245 and 254.

directly expressive$_1$ of the speakers mental state, the *vehicle* of that act is a sentence that expresses$_3$ a proposition.[87]

In the case of avowals, "I believe that it will rain" for example, the act of avowing is expressive$_1$ of the speaker's mental state (the belief that it will rain), while the vehicle of the avowal is a sentence that semantically expresses$_3$ *that the speaker is in that mental state* (the proposition that the speaker believes that that it will rain). In the case of assertion, "it will rain" for example, the act of asserting is expressive$_1$ of the speaker's *belief that it will rain*, while the vehicle of the assertion is a sentence that expresses$_3$ the proposition *that it will rain*. In the case of apology, "I apologize for what I did" for example, the act is expressive$_1$ of the speaker's *regret for what he did*, while the vehicle is a sentence that expresses$_3$ the proposition *that the speaker is apologizing.* And so forth *mutatis mutandis*.

The debts to Austin and Sellars here are apparent, but significant progress has been made. What Bar-On does is synthesize their two distinctions in a powerful way. From Austin we have the distinction between speech act and sentence, and from Sellars we have the distinction between act expression and semantic expression. Bar-On unites these insights in order to distinguish the phenomenon of *that which is action-expressed$_1$ by the performing of a speech act*, from *that which is semantically expressed$_3$ by the sentence that is the vehicle of that speech act.*[88] Importantly, her view supports a robust distinction between force and content: "Utterances that involve one and the same sentence, semantically individuated, may express$_1$ different types of conditions of the utterers, and

[87] Though Bar-On sometimes uses the term 'product' as well, I will stick with 'vehicle' for the sake of clarity. See Bar-On (2004), p. 255.

[88] Bar-On (2004), p. 259.

utterances involving different (semantically individuated) sentences may express$_1$ the same type of condition."[89]

Bar-On's Neo-Expressivist theory naturally suggests a particular approach to the question of hybrid metanormative views, which she has recently developed with the help of Matthew Chrisman and James Sias. I will return to that proposal at the end of this chapter, once I have finished giving the details of my own view, in order to clearly distinguish them. Before I do so, an *explanation* is needed for how and why action expression$_1$ takes place in the context of speech acts. The most complete, promising, and friendly such explanation on offer comes from the work of Mitch Green.

Section 3: Green on Non-Conventional Non-Conversational Implicature

In his book *Self-Expression*, Green offers a theory of speech acts and how they are able to express inner states. This theory is built on the central concepts of *showing*, *signaling*, and *speaker meaning*. To begin, Green observes that *showing* comes in three varieties:

> **Showing *that*:** providing compelling evidence for a conclusion in such a way as to make propositional knowledge available to an audience.
>
> **Showing *α*:** making some object or state of affairs *α* perceptible to an audience.
>
> **Showing *how*:** demonstrating what something is like in such a way as to make qualitative knowledge or empathy available to an audience.

[89] Bar-On (2004), p. 300. For a defense of this distinction, see Green (Forthcoming).

I could show *that* $a^2+b^2=c^2$ (where a, b, and c are the three sides of a right triangle) by writing out a proof in front of the class. It should be noted that this counts as *showing that* even if the students aren't paying close enough attention to acquire the knowledge of this theorem from my evidence. It is sufficient for showing *that* that I have made the evidence (and possibility of knowledge) available to the students in a way that they *could* grasp. But the audience at least has to have the *capacity* to understand my evidence (if there is a horse in the classroom, I have not shown *him* that $a^2+b^2=c^2$).[90]

Alternatively, I could show α–where α is my bruise–by making my bruise visible to you, or I could show *a rough texture* by getting you to feel it. What I am able to show α to an audience depends on the perceptual faculties of that audience together with features of our environment. So I could not show you the mice in a field 200 yards below us from our vantage point in an airplane, though perhaps I could show the mice to the eagle that is with us. Lastly, I could show *how* a skunk smells by applying friction to a scratch-and-sniff picture of a skunk, or I could show *how* trepidation feels by the quavering of my voice and the trembling in my hands. What I am able to show *how* to an audience depends on their perceptual and/or empathetic capacities (perceptual capacities for qualitative knowledge, empathetic capacities for sympathetic knowledge).[91]

There are three further important features of showing. First, these three species of showing often come bundled together, as when a slowly approaching group of clouds shows me *that* a storm is coming, α (the group of clouds), and also *how* a cumulonimbus

---

[90] Green (2007), p. 47.

[91] Green (2007), pp. 47–48. Though there may be an interesting connection between "knowing how" and the kind of phenomenal experience involved in "showing how," it is not *assumed* that they are related.

cloud looks. However, because it is also possible to show in only one of the three ways without showing in the other two, it is perspicuous to keep them conceptually distinct. Second, 'showing' is a success term. While one can *purport to show* (or "indicate") that something is or feels some way that it actually isn't or doesn't; one can only *show* what is really the case: ways things really are (*that*), states of affairs that obtain (α), and ways things really feel (*how*). Third, while showing can occur in all kinds of situations, specific features or processes may be *designed* to show.

This last observation leads us to another of Green's central concepts, that of a signal:

> **Signal:** Any feature of an entity that conveys information (including misinformation) and that was designed for its ability to convey that information.[92]

The notion of design invoked here may, but need not, be the result of intentional behavior. 'Design' is meant to encompass the work of intelligent agents, evolution by natural selection, and evolution by artificial selection. Using this definition of a signal, Green defines communication as follows:

> **Communication:** A signal's successfully conveying the information it was designed to convey.[93]

How do signaling and communication relate to showing? According to Green, whenever the cost of producing a signal is sufficiently high, signaling not only communicates the information that it was designed to convey, but it also *shows* that information. This is

[92] Green (2007), pp. 48–49.
[93] Green (2007), pp. 49.

because a signal's being a handicap (having a high cost) is one way of ensuring its reliability; it discourages *free-riding*.[94]

For example, a brightly colored tree frog may *signal* that he is poisonous, because his bright colors were naturally selected for their ability to convey this information. However, within a system in which poisonous tree frogs regularly signal in this way, the relative low cost of this signal makes cheating easy—hence the potential for brightly colored tree frogs that are not poisonous. A free-riding tree frog would not *show* that he is poisonous (remember, showing is a success term!) and neither does the actually poisonous tree frog (just being the color you are has a low cost), though they both signal it. In contrast, the extravagant train of a male peacock not only signals that he is viable, it *shows* his viability: he is able to survive despite carrying so much excess weight. Thus handicaps, by discouraging free-riding, constitute reliable/stable signals and consequently enable knowledge in their audience.[95] For our purposes here, the importance of showing (by reliable signaling) lies in its ability to explain *self-expression* generally, and *speaker meaning* in particular, via the mechanism of *making manifest*.

One aspect of speaker meaning that often goes unappreciated in the literature is the fact that there seem to be cases where we can *speaker-mean* without intending to produce an effect on an audience. For example, can't I speaker-mean something to an infant, or

---

[94] I do not mean to suggest that handicaps are the only means by which signals are capable of showing, just that it is the phenomena that will be most relevant to our discussion. See Green (2007), pp. 50–51, for descriptions of the other two methods (preference ordering and indices).

[95] Green (2007), pp. 51–52. Of course, I do not mean to suggest that peahens are capable of knowledge (though I won't rule it out either).

even to a crowd of people that I know are convinced I am a liar?[96] According to Green, what is central to speaker meaning is not that a speaker intends to produce an effect on an audience (though they often do), but rather that a speaker intends to make something *manifest*. That is, the speaker intends to make his or her intentional state "publicly accessible, but not necessarily in fact discerned by anyone….[the speaker] need only intend that the intentional state be 'out there': there to be discerned by anyone concerned to look whether or not anyone ever does."[97] This insight will return to play an important role in the argument of the next chapter.

We are now in a position to present Green's definition of illocutionary speaker meaning:

> *Illocutionary Speaker Meaning*: S illocutionarily speaker-means that P φ'ly, where φ is an illocutionary force, iff
>
> 1. S performs an action A intending that
>
> 2. in performing A, it be manifest that S is committed to P under force φ, and that it be manifest that S intends that (2).[98]

Illocutionary speaker meaning is communicated via speech acts, where Green says "a speech act is any act that can be performed by, under the right circumstances, speaker-

---

[96] Green (2007), pp. 60–61. Borge (2013) argues that most parents do not typically speak to infants in this way. However, this does not threaten Green's thesis that one coherently *can* so speak.

[97] Green (2007), p. 65. Another example might be speaker meaning by writing in a journal: one may speaker-mean what one writes, without ever expecting the journal to be read. Social media (Twitter, Facebook, etc.) has added a whole new dimension to this phenomenon. When a speaker posts online, she often only intends to put her message "out there" with little or no expectation that it will be read or produce an effect.

[98] Green (2007), p. 74. For a response to regress worries and definitions of the other two kinds of speaker meaning, see Green (2007) pp. 66-68.

meaning that one is doing so."[99] Green's theory of speech acts is in the same tradition of Austin, Sellars, and Bar-On that we have been tracing above. What Green adds to our story is an explicit definition of speaker meaning as distinct from force, semantic content, and action expression. And it is important to recognize that speaker meaning can come apart from all these things—just take the case of sarcasm. Suppose the two of us look out the window at a large approaching thunderstorm. The clouds are thick and dark and the rain is almost upon us. I have a habit for questioning the obvious (which you often find irritating), so it comes as no surprise when I ask "is it going to be a wet day?" You roll your eyes and reply by saying "no, it's going to be as dry as the Sahara desert!" What is going on here?

Green's theory allows us to parse the situation out in a clear and useful way. You have performed the action of uttering the sentence "no, it's going to be as dry as the Sahara desert" intending by your performance to make it manifest that you are committed to the proposition that it will be a very *wet* day indeed, under the force of assertion (of course, your intention so to make manifest is itself *also* made manifest). Here you have speaker-meant that it will be a very *wet* day, by using the vehicle of a sentence that semantically expresses$_3$ the proposition that it will be a *dry* day (a day as dry as an average day in Sahara desert, to be exact). Because you have speaker meant this with the force of an assertion, you also action express$_1$ your *belief* that it will be a *wet* day.[100]

---

[99] Green (2007), p. 70. These speech acts may, under appropriate conditions, be performed without including the performative in the utterance (e.g. "It is raining" may be uttered with assertive force in place of "I assert that it is raining").

[100] Of course, given the context you have certainly done more than this as well. You have pretended to disagree with me, you are protesting (or teasing) my habit of asking questions to which I should already know the answer, and you may even be expressing your irritation. Though this would also count as action expression$_1$, it is probably best analyzed as a form of conversational implicature, because it could easily be

Often what we speaker-mean diverges from what the sentences we use semantically express$_3$. Irony, metaphor, hyperbole, idioms—we have many tools in our literary chests for speaker-meaning something other than that which our sentences literally express$_3$. However, in the current discussion my range of target phenomena is somewhat narrower, and thus I will do my best to restrict the rest of our discussion in this chapter to cases in which speaker-meaning and semantic meaning run together more closely. Specifically, we began our foray into Green's work with the goal of elucidating *action expression$_1$*, and the unique way that speaker meaning (with particular speech act forces) allows us to make use of act expression$_1$ to show to each other our inner states.

As we saw above, many speech acts have an action expressive$_1$ dimension: assertions allow speakers to express$_1$ their beliefs, promises allow speakers to express$_1$ their intentions, apologies allow speakers to express$_1$ their regrets, etc. How is it that some speech act types enable speakers to action express$_1$ in this way? According to Green, it is ultimately by the same mechanism that the high cost of biological handicaps enables organisms who have them to show what their handicaps were designed to communicate. The *cost* of making a certain speech act can enable speakers who perform them to show what that speech act was designed to communicate.

It might at first sound strange to say that making a speech act can be *costly*. After all, biologically speaking, most speech acts burn very few calories.[101] Here it is helpful to distinguish *speech acts* from *acts of speech*. Acts of speech, the mere uttering of a string of words (as when testing a microphone on stage or trying to remember a phone number

---

cancelled ("No, I'm not irritated, I'm just giving you a hard time!"). More on implicature to come.

[101] Moon and Lindblom (2003).

while searching for a pen), are cheap indeed. If an act of speech happens to involve a properly formed sentence it can semantically express$_3$ a proposition, but that is the end of it. Mere acts of speech involve no speaker meaning or act expression$_1$. In contrast, *speech acts* are governed by rules that determine the risks one undertakes by making them—these are Austin's "infelicity conditions" that we discussed above. Green explains the connection between speech act rules and expression$_1$ in the following way: because some speech acts are defined by sets of norms that commit the speaker to costly enough risks, they can have an expressive$_1$ dimension.

Speech acts are social artifacts, and this particular class of *action expressive$_1$ speech acts* involves using costly signals that have been designed over time to enable speakers to express what they do by a process of cultural selection. Just like Bar-On's Neo-Expressivism above, Green understands the notion of *action expression$_1$* to be a species of *showing*—a success term. When an expressive speech act is performed sincerely, the speaker enables his audience to know his internal states. For example, when I sincerely utter the sentence "it is raining" with assertoric force, I simultaneously (1) express$_3$ the proposition *that it is raining* via the semantic and syntactic properties of the sentence that I have uttered, (2) speaker-mean *that it is raining* by performing this action with the force of an assertion, and (3) express$_1$ *that I believe that it is raining* by showing that I have this belief. The illocutionary force with which a speech act is performed determines the norms by which it is judged, the norms determine the category of risk that the speaker undertakes in the performance, and type of risk in turn determines the intentional states that the speaker thereby communicates (and *expresses$_1$* if he is sincere).[102]

[102] Green (2007), p. 72.

The expressive dimension of speech acts enables us to mean more than we say, and hence is a good candidate for being an instance of *implicature*.[103] In the last chapter, we discussed metanormative views that appeal broadly to the phenomena of conversational and conventional implicature. These are the types of implicature that Grice spends most of his time discussing. However, he does also briefly mention a "non-conventional non-conversational" species of implicature.[104] According to Green, this is the mechanism by which speakers express$_1$ their inner states when they perform certain expressive speech acts like asserting or promising. To better appreciate what is distinctive of non-conventional non-conversational implicature, let's look more carefully at the way that speakers express$_1$ their beliefs by making assertions.

When I utter the sentence "it is raining" with the force of an assertion, I express$_1$ my *belief that it is raining* in a way that is not cancellable in the Gricean sense that we discussed in the last chapter. That is to say, if I follow up my assertion with an explicit denial of the belief that I expressed$_1$ ("but I don't believe that it is raining") I will produce confusion and puzzlement in my audience. This is the widely recognized phenomenon of Moorean absurdity. The fact that the expression$_1$ generated by speakers who perform speech acts is not cancellable in this way is a good indication that it is not an instance of conversational implicature.

It also does not survive embedding. Because the expression$_1$ of belief is generated in virtue of the fact that I am understood to be performing an assertion, if the assertive force

---

[103] By 'meaning' here I mean something akin to, though not necessarily identical to 'speaker meaning.' Identifying the category of meaning into which implicature falls is a significant task, and I intend to be neutral about it here.

[104] Grice (1989) p. 28.

is not present then neither is the expression$_1$ of belief. As Bar-On observes, embedding statements in conditionals, propositional attitude ascriptions, indirect quotations, negations, etc. constitutes a "force stripping context."[105] No assertive force, no expression$_1$ of belief. Likewise, uttering a sentence with another sort of *non-assertive* force—like that of a *question* or a *wild guess*—is another good way to ensure that one does not express$_1$ that one believes what one says. On top of all this, there is the obvious fact that the expression is not generated in virtue of the conventional meaning of any particular term that composes the vehicle (sentence) of the speech act. In fact, as we have already seen, the very same sentence may be used by a speaker to perform different speech acts at different times in different contexts, and thereby to express$_1$ different mental states. These considerations together constitute a clear indication that when expression$_1$ is generated by a speech act, it is not an instance of conventional implicature.

To sum up: the expression$_1$ of mental states generated by speakers in virtue of their performing particular speech acts is not dependent on the conventional meanings of the terms the speaker employs (otherwise they would survive *embedding*), neither is it dependent on the general pragmatic conversational principles and maxims (otherwise they would be *explicitly cancellable*).[106] Thus Green concludes, if this form of expression$_1$ is a form of implicature—a way of meaning more than we say—but is neither conventionally nor conversationally analyzable, it must be an instance of Grice's elusive "non-conventional non-conversational" category of implicature (or "NNI").

---

[105] Bar-On (2004), p. 233–234.

[106] There may yet be a temptation to explain speech acts by appeal to conventions. This temptation comes from the tradition of *force conventionalism* stretching back to Austin (1962) and Searle (1969), and resisted by Strawson (1964). For a helpful discussion of the relevant aspects of this debate, see Green (2015).

When a speaker performs an expressive$_1$ speech act sincerely, the resulting implicature is an instance of *showing*. But what if the speaker is not *in* the mental state that is typically expressed by speech acts of type she is performing? As noted above, 'expression$_1$'—being a species of showing—is a success term. So, for example, an insincere asserter *by definition* cannot express$_1$ a belief that she does not have. Nevertheless, the speaker still *implicates* that she believes what she says—that is what makes her lie a successful instance of lying. How is this possible?

The explanation is this: when a speaker performs a speech act insincerely, in order for it to count as an instance of the speech act type that it does, her performance must still be understood to be subject to evaluation according to the constitutive norms of that speech act type. Thus, she still undertakes the same risks, and thereby still communicates what the speech act was designed to communicate. Because the intentional state in question is not there to be shown, insincere speakers *purport to show* their intentional states instead of *actually showing* them. This "purporting" still counts as a non-conventional non-conversational implicature. In other words, the speaker still *means* and therefore *communicates* more than she says even though her lack of sincerity causes her communicative act to fall short of *showing* her inner states (which are not there to be shown). Without showing, there is no genuine *expression$_1$*.[107] Some tables representing the possibilities here may be helpful.

[107] See Green (2007), pp. 99–104. Green argues that NNI's may be produced by cases of self-expression beyond just speech acts. The phenomena of NNI is explained by the mechanism of reliable signals (handicaps) as they contribute to speaker-meaning, and hence can be generated in any instance of self-expression. While I agree with this, the topic of this paper is restricted to speech acts.

Here is a table outlining an example of an assertion:

			Speaker is Sincere	**Speaker is Insincere**
What is *Communicated:*	What is *Said*:		*P*: "It is raining"	
	What is *Implicated* (NNI)…	…via *Expressing*:	"S believes that *P*"	–
		…via *Purporting*:	–	"S believes that *P*"
What is *Risked:*			Loss of credibility if ¬(S believes that *P*)	

Table 1: Speaker utters "It is raining" with assertoric force.[108]

And here is a table outlining an example of promising:

			Speaker is Sincere	**Speaker is Insincere**
What is *Communicated:*	What is *Said*:		*P*: "S will meet A for lunch Tuesday"	
	What is *Implicated* (NNI)…	…via *Expressing*:	"S intends that *P*"	–
		…via *Purporting*:	–	"S intends that *P*"
What is *Risked:*			Loss of reputation if ¬(S intends that *P*)	

Table 2: Speaker utters "I will meet you for lunch Tuesday" with promissory force.

According to Green, what is *communicated* is the same whether the speaker is sincere or insincere. This is because whether sincere or insincere, the speaker still utters the sentence in question with the force of whichever speech act she is performing, governed

[108] I should also briefly note that the investigation of the norms of assertion is a hotly contested debate in epistemology. While I readily admit that this table is incomplete, my discussion should make clear how particular theories of assertion would complete it, and the argument of this paper is neutral with respect to the different ways of doing so. See Williamson (1996), and Brown & Cappelen (2011).

by its constitutive felicity conditions, and thereby undertakes the same risk. This is why what is *implicated* remains the same, though whether the implicature is generated due to the speaker's *expressing*$_1$ or only *purporting to express* depends on whether the speaker in fact has the belief. To be clear, a speaker may generate other implicatures by using particular words (conventional implicatures) or by performing her speech act in particular contexts (conversational implicatures). But as long as we also hold fixed the risks and norms constitutive of the speech act type being performed, her NNI will remain constant.

We are now in a position to say something about David Copp's distinction between implicatures and *simplicatures*. Copp makes this distinction because he takes implicatures to be communicated intentionally, which (in Green's terms) is to say they are always a part of what is speaker-meant. In contrast, Copp thinks that our target phenomenon here is actually part of a broader category, one that also includes information beyond what is said that is communicated unintentionally. I think this is an unhelpfully broad category, as it can include all sorts of information that may be conveyed by an utterance. When I say "it is raining" under ordinary circumstances I convey to you the information that I know how to speak English, but clearly this sense of communication is far afield from that which we are interested in.

At the same time, I will admit that we are not in *every* instance of speech act performance explicitly intending to communicate the information that is action expressed$_1$ by our speech act. I do not necessarily form the conscious occurrent intention to convey to you that I believe that it is raining every time I assert to you "it is raining." In fact, most of the time I probably only explicitly intend to tell you *that it is raining*. So I will agree that speech act action expression$_1$ is not always a component of speaker meaning. Nevertheless,

I think it is more useful to broaden our category of implicature to include these cases, than it is to open ourselves up to the free-for-all that is simplicature, including as it does all manner of natural meaning, causal expression$_2$, etc. At the very least, speakers *do* intend to perform a speech act that is of a type constituted by certain norms, even if they don't always have the implications of those norms consciously in view. Thus I find it more useful to describe speech act expression$_1$ as an instance of non-conversational non-conventional implicature, arising as it does from the risks undertaken by speakers in performing the speech acts of the types they do, regardless of whether that expresssion$_1$ is a component of speaker meaning in a given instance or not.

Part III: The Positive View of Normative Language

Section 1: The Hybrid Speech Act Theory

Towards the beginning of *How to Do Things with Words* Austin makes the following remark:

> It has come to be commonly held that many utterances which look like statements are either not intended at all, or only intended in part, to record or impart straightforward information about the facts: for example, 'ethical propositions' are perhaps intended, solely or partly, to evince emotion or to prescribe conduct or to influence it in special ways.[109]

It is Austin's passing caveat, "or partly," that is to be our inspiration in this section. In this section, my goal is to bring the resources from Part II of this chapter to bear in making

[109] Austin (1962), pp. 2–3.

precise what Austin describes as the "commonly held" view of normative judgments. To move just a bit nearer to the target, consider an observation from Bar-On:

> Unless there are independent reasons to think that one cannot express$_1$ two distinct states with one bit of intentional behavior, we should perhaps remain neutral on the matter….And I see no general reason to maintain that one cannot express$_1$ more than one state with a bit of behavior.[110]

The first step towards my positive view is a straightforward application of the spirit of the hybrid metanormative approach: when speakers make normative claims, they express both a cognitive state (like a belief, opinion, etc.) and a moving state (like a desire, emotion, affective reaction, etc.). Here is the second step: the notion of "expression" at play here is that of action expression$_1$. Now consider a possible hybrid view that David Copp mentions in passing:

> In some cases "expression" refers to the relation between a person performing a speech act and the psychological state having which is the 'sincerity condition' of that speech act. In this sense, for example, an apology can be said to express regret and an assertion to express a belief. Of course, a person can assert something without believing it. A person who asserts that *p* expresses the belief that *p* in that, roughly, if she is sincere in what she says, and if she knows her own mind and knows the meaning of the sentence she utters, she believes that *p*. More generally, a person performing a speech act thereby "sincerity-expresses" a state of mind just in case, if she is sincere in what she says, and if she knows her own mind and knows the meaning of the sentence she utters, she is in that state of mind.[111]

The best I can make out is that Copp is attempting to talk about the phenomenon of speech act expression$_1$ that Green so carefully analyzes, explains, and diagnoses as non-

[110] Bar-On (2004), pp. 306–307.
[111] Copp (2009), p. 173.

conversational non-conventional implicature. With Bar-On and Green's theories in mind, let's clean up the proposal a bit.

I believe that a more precise way to put what Copp is trying to say would go something like this: Speakers who make normative claims are performing speech acts, the felicity conditions of which include the requirements that the speaker be in both cognitive and moving states of mind. This is because these speech acts are "designed" (in both Green's and Austin's sense) to be used by "persons having certain thoughts and feelings."[112] Because the speech acts in question have the felicity conditions they do, speakers making use of them undertake the social risk of being criticized if they are found out as not being in the right states of mind. Hence speakers express$_1$ those cognitive and moving states by *showing* them when they are fully sincere, and by *purporting to express* them even when insincere. Either way, a non-conventional non-conversational implicature is present to the effect that the speaker both believes and is motivated/affected in a way harmonious with the normative claim she is making.

Copp dismisses this "sincerity expression" strategy for the following reason:

> Suppose Anna says capital punishment is morally wrong. I think she could know her own mind, know the meaning of the sentence she utters, and be sincere in asserting that capital punishment is wrong, even if she does not actually disapprove of capital punishment. If so, she does not sincerity-express disapproval. Perhaps, as Michael Smith suggests, a failure to disapprove of capital punishment would "cast serious doubt" on her sincerity (1994: 7), but this does not mean that she must actually be insincere. There need not be any deceit or pretense on her part. She might be convinced capital punishment is wrong and be honestly expressing her belief it is wrong even if her feelings are at odds with her belief. She might agree that it is perverse of her not to feel disapproval. So I think her assertion need not be insincere. If she does express disapproval in saying capital

[112] The reader should recognize the quote from Austin (1962), p. 15. However, it is the same idea appealed to in Green (2007), for example on pp. 72–73.

> punishment is wrong, it seems that this is not a matter of sincerity-expression.[113]

After the careful discussion from Part II, we are now in a position to see where Copp goes wrong in this assessment. First, notice that Copp speaks of Anna's speech act as being, essentially, an *assertion with normative content*. This confuses the issue. Remember, speech acts are individuated by their felicity conditions. True, the speech act of assertion has as one of its felicity conditions that speakers performing it must *believe* what they say. However, it does not have has one of its felicity conditions *that speakers must be motivated in a way congruous with what they are saying*. In that sense Copp is totally right: a speaker may *assert* that capital punishment is wrong without disapproving of capital punishment, and count as fully sincere, as long as she believes what she says. This just falls out of the accepted definition of assertion.[114]

Copp's objection misses the point precisely because he does not take his own "sincerity expression" proposal seriously. To take the proposal seriously is to envision a speech act *with felicity conditions that distinguish it from assertion proper*. That is to say, were there a speech act such that it was "designed for use" by persons with certain cognitive *and* moving mental states, then that speech act would (by definition) *simply not be* the speech act of assertion. Rather, because it would contain in its felicity conditions the requirements that speakers both *believe* and be *moved* (in either the motivational or affective sense) by what they say *it would constitute a distinct speech act type*. This is the

---

[113] Copp (2009) pp. 173–174.

[114] Whether it is psychologically *possible* for a speaker to do this is up for debate. Some Kantians will say no. I believe that Humeans are require to say yes. We will discuss more about what is going on *inside* normative judges in the next chapter.

approach that I take. After describing the view in broad strokes, we will return to Copp's example of Anna.

My own positive view is called the Hybrid Speech Act Theory. It consists in taking seriously the common-sense spirit of the hybrid metanormative approach, and combining it with our best understanding of speech acts and action expression$_1$. When we do this, the natural response is to posit as many different distinct speech acts available as there are legitimately distinct normative domains. Each distinct normative domain—epistemic, prudential, ethical, aesthetic, rational, etc.—has arisen with somewhat different aims, and consequently provides us with a somewhat different species of speech act, distinguished by its own unique felicity conditions. What these various normative speech acts have in common is that they all require of speakers the dual possession of both cognitive and moving states—this is the insight of the hybrid metanormative approach. Where the different normative speech acts differ is in the precise nature of the mental states speakers performing them are required to have.

Perhaps some normative domains arose, in part, for us to show our affective feelings or emotions to one another. Perhaps other domains arose for us to express our intentions, motivations, and plans. Still other domains might have arisen for us to give recommendations or commands, or to show each other the dispositions we aspire to when thinking most clearly. I have not yet committed myself to the specific content of the respective felicity conditions of the various speech acts in the different normative domains, as long as they "have it both ways" in truly hybrid fashion. It may turn out that we have more or fewer distinct normative domains than we thought we had. For example, we may decide that there is a difference between the *moral* and the *ethical* in Bernard Williams'

senses of the terms.[115] Or we may discover that the rational/prudential distinction collapses under scrutiny. Whatever the case may be, my hybrid speech framework can help us to think clearly about what is at stake. In fact, we now have an additional way to adjudicate disagreements about whether two normative domains are truly distinct: take a look at whether claims made in the two domains are governed by different felicity conditions.

Now let's return to Copp's example of Anna. Perhaps a clarified form of the objection still applies: a speaker may *say* that capital punishment is wrong without feeling any disapproval towards capital punishment, and with no deceit or pretense on her part. In order for this true observation to count as an objection to the hybrid speech act view, it would require the mistaken assumption that failing felicity conditions requires deceit or pretense. As discussed in Part II above, we do not generally believe that this is the case. After all, a speaker may fail the felicity conditions of assertion without any intended deceit, simply by lacking the self-awareness to recognize that she does not believe what she says. So then, in order to clearly describe the scenario we need to carefully note what sort of speech act Anna is attempting to perform (and hence what felicity conditions apply to her). To see whether it is possible for Copp's example to create a problem for my view, let's examine the two possible ways it could play out:

(1) Anna says "capital punishment is wrong" meaning this as an *assertion*. She believes what she says, so she counts as fully sincere. She expresses$_1$ her belief by the same mechanism of non-conversational non-conventional implicature generated by all assertions. She feels no disapproval of capital punishment, but this in no way undermines the sincerity of her assertion, as it was never required.

(2) Anna says "capital punishment is wrong" meaning this as a *moral speech act*. She believes what she says, so she counts as partially sincere. Moral

[115] Williams (1985), pp. 6–7.

> speech acts require more than just belief, however, they also require the appropriate motivation or feeling. Anna does not feel disapproval of capital punishment, so to that extent her speech act is infelicitous. However, she is not intending to be deceitful in any way, and in fact recognizes that it is perverse of her not to feel disapproval. She is troubled by the fact that her feelings are at odds with her belief.

Both (1) and (2) are perfectly consistent with my view. Of course, we will want to know why a speaker would choose to merely assert a claim with moral content rather than perform the moral speech act. I will return to this question in section 3 below.

While recognizing the possibility of circumstances in which (1) occurs, I believe that that our default assumption about speakers who utter normative claims (without qualification) is more like (2). That is to say, speakers uttering unqualified normative sentences in everyday circumstances *are not ordinarily taken to be performing assertions*. Rather, they are taken to be performing a related but distinct *normative* speech act. These normative speech acts, like assertions, require as part of their felicity conditions that speakers believe what they are saying. However, they also have an additional requirement for full felicity: that speakers possess some measure of motivation, affection, or feeling appropriate to the claim they are making. Thus normative speech acts are—to adopt Austin's way of speaking—accepted conventional procedures designed for use by persons having certain thoughts *and* feelings, *and* for the inauguration of certain consequential conduct on the part of the participants.

The hybrid speech act theory posits different associated speech acts within the different domains of normative speech—speech acts which speakers perform to express their cognitive and moving states. These different speech acts have different constitutive felicity conditions, enabling us to express different clusters of motivational or affective

states depending on which speech act we are performing. If I make the aesthetic judgment, "Wells Cathedral is a stunning example of Early English Gothic style—a style which is particularly beautiful," to understand what I have done you must understand me as binding myself by certain norms. If I follow my statement up with "…and I really don't like the way it looks" you'd be puzzled and confused in a way similar to the confusion you would feel were I to say "…and I don't believe it." Similarly, were I to make the ethical claim "environmental apathy is a vice that is destroying our planet" my audience would understand me to have submitted myself to a certain set of norms. In doing so, the stakes of our conversation have beyond the norms of assertion and into the norms of ethical judgment. If I follow up my claim with "…and I don't care at all whether I myself develop such a vice," my audience will be rightly puzzled.

According to the hybrid speech act view, aesthetic judgment, ethical judgment, epistemic judgment, prudential judgment, etc. pick out different members of a family of speech acts with overlapping—but nevertheless distinct—constitutive norms. It is a hybrid view in that each of these speech acts is taken to require in its constitutive norms the possession of both cognitive and motivational/affective states, though the precise nature of these states differs from one normative domain to the other. And the hybrid speech act view makes sense of why this would be the case. Each of these normative domains of discourse has developed as a way for us to solve practical problems, to determine what it is that we should do, to coordinate our actions with each other, to determine what we believe and desire, and/or to show each other our inner states.

The best way to get a grip on this view is through examples. Here is a table outlining a moral speech act in the same way we outlined assertions and promises above:

			Speaker is Sincere	**Speaker is Not Sincere**
What is *Communicated:*	What is *Said*:		*P*: "Environmental apathy is a vice"	
	What is *Implicated* (NNI)…	…via *Expressing*:	1. "S believes that *P*" 2. "S is motivated in accordance with *P*"	—
		…via *Purporting*:	—	1. "S believes that *P*" 2. "S is motivated in accordance with *P*"
What is *Risked:*			1. Loss of credibility if ¬(S believes that *P*) 2. Loss of reputation if ¬(S is motivated *P*'ly)	

Table 3: Speaker utters "environmental apathy is a vice" with moral judgment force.

And here is an example of an aesthetic speech act:

			Speaker is Sincere	**Speaker is Not Sincere**
What is *Communicated:*	What is *Said*:		*P*: "Wells Cathedral is beautiful"	
	What is *Implicated* (NNI)…	…via *Expressing*:	1. "S believes that *P*" 2. "S is affected in accordance with *P*"	—
		…via *Purporting*:	—	1. "S believes that *P*" 2. "S is affected in accordance with *P*"
What is *Risked:*			1. Loss of credibility if ¬(S believes that *P*) 2. Loss of reputation[116] if ¬(S is affected *P*'ly)	

Table 4: Speaker utters "Wells Cathedral is beautiful" with aesthetic judgment force.

[116] Though I use the same term—'reputation'—in the examples of what is risked by a moral judge and what is risked by an aesthetic judge, it should be clear that the manner in which one puts one's reputation on the line can vary significantly from one normative domain to another. Accusations of moral "hypocrisy" carry with them very different (and probably more serious) social costs than do accusations of being a "pretender" to aesthetic experiences.

The hybrid speech act view consists in the claim that such analyses may be given of all normative claims made in the context of public speech. For each distinct normative domain, there is a corresponding distinct table outlining what is risked (derived from the unique felicity conditions of the normative speech act type), that may be further precisified by filling it out with some particular claim or other from within that domain. To reiterate a point from above: I am not necessarily committed to a specific answer to the question of what the felicity conditions *are precisely* for each of these speech acts in their respective domains. These are plausible, but revisable examples.

Normative speech acts are designed for use by speakers in at least two different mental states, one cognitive and the other moving. It seems plausible to me that in the case of moral judgment these states are belief and motivation, while in the case of aesthetic judgment these states are belief and affect. I will provide additional reasons for thinking so in Sections 3 and 4 respectively. Importantly, nothing in the Hybrid Speech Act view commits one to the idea that the cognitive and moving elements must necessarily come together.

On the contrary, we have already seen that remaining consistent with Psychological Humeanism requires allowing that either one could be had without the other. Thus speakers could fail to be fully sincere either by lacking the motivational/affective profile harmonious with their beliefs, or vice versa. Such a falling short of wholehearted sincerity would render their unqualified normative claims misleading, at least in part, even if not intentionally deceptive. Later on we will see how this feature of HSA allows it to affirm motivational internalism while doing full justice to externalist intuitions.

Section 2: An Argument for Positing Novel Speech Act Types

The Hybrid Speech Act view makes a bold claim. After all, HSA implies that Austin's and Searle's—already quite extensive—catalogues of speech act types are still incomplete. Why should we suppose that there exists a whole class of hitherto unrecognized speech acts? Though the hybrid speech act view gets considerable support from its unique ability to "have it both ways" as a hybrid metanormative theory, this on its own may not convince philosophers of language satisfied with the status quo in speech act theory. An independent argument here will be helpful, so I will take a moment to discuss the following question: "What sort of considerations should justify us in positing previously unrecognized types of speech acts?"

Surprisingly, very little has been written on this question. One of the clearest examples comes from Richard Gale's 1971 essay "The Fictive Use of Language." In this essay, Gale is interested in explaining how a speaker's uttering declarative sentences in the course of creating a fiction need not involve taking on ontological commitments. For my purposes here I need not take a stand on whether Gale's analysis of the "fictive use" in particular is ultimately correct. However, it will be instructive for us to excavate the method that Gale implicitly uses, because his conclusion is that the fictive use is *itself* a previously unrecognized type of speech act. In that sense his argument is a cousin to the one I am making. Once I have exposed and reanimated Gale's methodology, I'll argue in favor of it and rebrand it using resources from the beginning of this chapter.

Gale's first observation about fictional language depends on Austin's distinction between locutionary and illocutionary acts. This is just another way of describing Green's distinction between acts of speech and speech acts discussed above. Merely uttering a

sentence is a locutionary act (an act of speech). *Asserting*, *promising*, *apologizing*, etc. are illocutionary acts (speech acts). Most of the time our illocutionary acts are at least partly constituted by locutionary acts, and most of the time our locutionary acts are used as vehicles to perform illocutionary acts. However they can come apart, as when I perform a locutionary act (but no illocutionary act) in the course of testing a microphone.[117]

With this distinction in place, Gale makes the following observation:

> The locutionary-illocutionary distinction can now be put to work. When a person makes a fictive use of a sentence he says exactly what he would say if he were to use it non-fictively. This is obvious because we do not need a special dictionary or grammar book in order to understand a fictive use of language. Words and sentences occurring in a fictive narration do not acquire a new meaning, nor do our ordinary syntactical rules cease to apply to such sentences.[118]

In other words, the very same locutionary act (act of speech) may be used to perform either a fictive or non-fictive illocutionary (speech act) use of language. This is the first phenomenon that we should be on the lookout for when deciding whether to posit a new speech act type. It will usually be the case that it can be performed by uttering sentences which may also be used to perform speech acts of other types. For example, Charles Dickens may utter the sentence "It was the best of times, it was the worst of times…" and thereby mean it as part of his fictional work *A Tale of Two Cities*. But I could utter this very same sentence and mean it as an *assertion* about the world (whether I have read

---

[117] Arguably, we may also be able to perform illocutionary acts without performing locutionary acts. For example, I may be able to *assert* to you that I have a broken arm without saying anything, by demonstratively thrusting my plaster cast before your face (in the right context).

[118] Gale (1971), p. 327.

Dickens or not). The very same sentence may be used with different types of speech act force.

The next observation Gale makes is that we do not criticize speakers who use language fictively when they say things that are literally false. Here is the most relevant part of his explanation:

> [T]o say that what a person said is false reflects adversely on this person, viz. that he spoke carelessly, without any grounds, or insincerely, etc. Since the person who uses a sentence fictively does not perform an assertive illocutionary act, the pragmatic implication of non-fictive use of this sentence, viz. that he believes what he says, is cancelled. Thus it would be unfair to charge him with saying something false. And since we would not be willing to charge him with saying something false, neither would we be willing to honour him with having said something true. But all of this is quite consistent with his saying things that are true or false.[119]

Gale's discussion here suggests a criterion for speech act-hood on which he is implicitly relying. Imagine that a speaker utters the same sentence in two different contexts, and we all agree that a criticism of infelicity would be appropriate in the one context but would not be appropriate in the other, *and* the only relevant difference between the contexts is the speaker's intention (of which difference the audience is aware). In such a case, Gale seems to think, we have reason to conclude that the speaker is performing two different speech acts. Specifically, criticisms that may be appropriately aimed at an asserter (e.g. "what you said was false!") may not be appropriately directed at a fictive speaker—even if they utter the same sentence.

Even in this short passage from Gale, we can already see hints of the picture that was laid out in full earlier in this chapter. For example, the connection between appropriate

---

[119] Gale (1971), p. 328.

criticism and felicity conditions, as well as the connection between performing an act that has certain felicity conditions and the communication of a "pragmatic implication" that one satisfies those conditions. Part of what is distinctive about the fictive speech act, according to Gale, "is that it consists in a desisting from performing any other illocutionary acts" and hence does not involve sticking one's neck out in the same sort of way that asserting does.[120]

At the same time, Gale thinks, the fictive speech act also has positive felicity conditions of its own. It is not merely defined by getting a speaker off the hook:

> A fictive illocutionary act, like any positive illocutionary act, can be unsuccessful. Even though it says 'Kraft Theatre' before the performance of the play on TV or a narration begins with 'once upon a time' it might fail to be the case that a fictive use of language occurs in what follows. The author of the 'fictional' play or story might have intended its major characters and incidents to be real. In such a case his fictive disclaimers, whether made explicitly or whether implicit in the context, are deceitful. Even were the author to preface his narration with 'I hereby tell you the fictional story that' it would not ensure that he would be making a fictive use of language in what follows.[121]

Thus on Gale's view part of what it takes for a fictive speech act to come off felicitously, is for the author to intend for the story's main characters and events *not* to be real. This seems plausible, but I would like to remind the reader that for our purposes in this chapter it does not matter whether Gale is entirely correct in his analysis of fictional discourse. What *does* matter is the fact that Gale takes the observation of distinct felicity conditions to be evidence of a novel speech act. This fits well with the view endorsed earlier in this

[120] Gale (1971), p. 335.
[121] Gale (1971), p. 336.

chapter—taken largely from Mitch Green's work in speech act theory—that speech acts are individuated by the norms that govern them.[122]

The final consideration Gale raises involves the fact that most speech acts are (or at least can be) introduced by an illocutionary indicator. If we wish to make the illocutionary force of our act explicit, we may do so by saying "I hereby assert" or "I hereby apologize." Gale points out that this might at first seem like a respect in which fictional uses of language are different from other illocutionary acts:

> A fictive use of language usually is not preceded by any explicit fictive illocutionary indicator to indicate that one is engaging in an act of illocutionary disengagement. The context of the speech act usually indicates that it involves a fictive use of language. E.g. it says Theatre' over the marquee or 'Novel' on the cover of the book, or one's narration is prefaced by the claim that any resemblance to persons living or dead is purely coincidental, etc. The closest thing in ordinary discourse to an explicit fictive indicator is 'Once upon a time.' 'I hereby ask you to imagine that,' although not ordinarily used as a way of introducing a fictional narration, can also serve as an explicit fictive illocutionary indicator.[123]

---

[122] This happy coincidence is perhaps not altogether the product of chance, considering that Richard Gale was one of Mitch Green's teachers during his time studying at the University of Pittsburgh.

[123] Gale (1971), p. 336. An interesting twist on this, and an apparent flouting of Gale's posited felicity conditions, may be seen in the Cohen Brother's *Fargo* franchise. Both the film and each episode of the television show begin with the words "THIS IS A TRUE STORY." Then follows a (purported) identification of the year and location of the events depicted. The preface always ends: "At the request of the survivors, the names have been changed. Out of respect for the dead, the rest has been told exactly as it occurred."

Given the context, Gale would say the preface itself is embedded in a fictional use, and hence is not truly infelicitous. The implicit directive is thus something like "[Imagine that] this is a true story"—which, arguably, puts the viewer in a somewhat different frame of mind than they otherwise would bring to the film. Perhaps it contributes to the atmosphere which the Cohen brothers are able to create of an earnest and simple (but sinister) Midwest American environment, and heightens the disturbing effect of the violence portrayed. An interesting question to investigate.

The central point of this passage is that in order for something to count as a speech act, indicators must exist whereby speakers can explicitly flag their utterances as having the force of that type. However, it need not be the case that speakers *usually do so*. There may be some sorts of speech acts for which context ordinarily supplies the indicator, and it would be unusual for a speaker to explicitly flag it (unless there were a significant chance that she would be misunderstood otherwise). Gale thinks that the fictive speech act is one of these speech acts that we usually do not find it necessary to preface with an illocutionary indicator. Plausibly, assertion itself is another; I do not usually find it necessary to begin my assertions with "I hereby assert..." But we *can* do so if we need to, and that is an important part of what it takes for something to count as a distinct speech act.

As promised, I will now provide some independent considerations in favor of these criteria on my own terms and from a somewhat different direction. I will also add a new criterion that Gale could easily have appealed to, though he did not do so explicitly. At the end of the day I am more confident in the criteria themselves than I am in Gale's claim that the fictive use satisfies them. I leave the status of his theory of fictional discourse an open question. However, I will fully take on board my clarified and fully specified version of the criteria; I believe that they are the best available at the moment. Thus this section will conclude with a brief discussion of how well the Hybrid Speech Act Theory fares against them.

To begin, imagine a linguistic community about whose speech act practices we have only partial knowledge. Suppose we know only that they possess analogues to our speech acts of questioning and asserting, and we must now discover what other speech act types they possess. What sorts of evidence or criteria would we look for, in order to

consider ourselves justified in positing additional speech acts to explain the behavior of this community? Specifically, try to imagine the sort of observations that could lead us to believe that they possess not only the speech act of assertion, but also the related (but distinct) speech act of offering one's opinion—of *opining*?

First observe that, if the community really does have the speech act of opining then it may be an act they can perform by uttering the *very same sentences* that they might use to perform assertions (e.g. sentences like "the Warriors are the more balanced team" or "this wine smells of honeysuckle"). Remember the independence of force and content: As Bar-On says "Utterances that involve one and the same sentence, semantically individuated, may [action] express different types of conditions of the utterers, and utterances involving different (semantically individuated) sentences may [action] express the same type of condition."[124] Another way to put this is that we need to keep in mind Green's distinction between speech acts and acts of speech (which is also more or less Austin's distinction between illocutionary and locutionary acts). I would describe this less as a criterion, and more an important background assumption about the nature of pragmatics.

Hence the first real sign that you have two distinct speech acts on your hands is that a speaker can use the *same* sentence to action express$_1$ *different* things by meaning their utterances with different forces. Because the *sentence* is the same in both situations, the difference in expression cannot be captured by appeal to the phenomenon of conventional implicature (or *simplicature*). If the expression is also not explicable in terms of general conversational maxims and principles, and thus is not a species of conversational

[124] Bar-On (2004), p. 300. For a defense of this distinction, see Green.

implicature, then this is a sign that it is generated by the assumed observance of speech act norms. The key here, I believe, is to identify distinct patterns in the sorts of ways that speakers can be held accountable for what they say.

The observation of a unique pattern of action expression$_1$ is strong evidence that the act satisfies the first real criterion, which is that a distinct set of norms may be identified *by which speakers performing the act are taken to be bound*. These felicity conditions serve to individuate speech act types, hence the moniker "constitutive norms." In our example, the speech acts of asserting and opining overlap in their constitutive norms—both require that a speaker believe what she is saying. However, asserting has an added justificatory requirement that opining does not. Suppose we observe patterns of behavior in which speakers are systematically criticized or challenged for not believing what they say, but are neither criticized nor challenged when the belief in what they say lacks justification. This would constitute evidence that the community in question has a *constative* act that is less epistemically demanding than full-fledged assertion. Because felicity conditions are *constitutive of* the speech act types, identifying a new, distinct, and stable pattern of ways a speaker may be criticized for their action performance is one of the best sources of evidence that a novel speech act should be posited.

Not every speech act type possesses a set of norms that enables speakers to action express$_1$. However, many do and they are the easiest to identify. This is one reason why I am hesitant to fully endorse Gale's fictive speech act, because it does not have an expressive dimension. In order to count as a distinct speech act type, it is essential that the act in question possess a unique set of felicity conditions. Given that action expression itself is not essential, it is possible that Gale is right. But without the added evidence of

unique felicity conditions that action expression$_1$ would provide, I am less confident in positing them—even in Gale's fictive case.

The second criterion for counting as a distinct speech act is that there is an accepted way for a speaker to explicitly signal that she is performing that particular speech act, but that she need not always make use of this signal. Thus we may begin our assertions with "I *assert*…" or our promises with "I hereby *promise*…" The intelligibility of such a signal is essential for something to count as a distinct speech act, even though it may be unusual for speakers to make use of it. Observing speakers who say "I hereby opine that…" would reinforce our justification for positing a new speech act of opining in the hypothetical case at hand.[125] Similarly, we might observe speakers to sometimes follow up their simple declarative utterances with "I didn't mean that as an assertion, I only meant to offer my opinion." This would also count in favor of an opining practice, as long as such cases do not confuse or puzzle the audience. Gale's discussion of this in the fictive case discussed above is very clear and convincing.

This last remark points the way towards an important piece of evidence that can count in favor of positing a new speech act type: observing a corresponding broader practice. This is a practice with its own set of internal norms and goals, within which the putative speech act is able to play an important functional role—a role played largely in virtue of the fact that it has the constitutive norms that it does. Sometimes the practice is

[125] Perhaps we possess more natural (though less explicit) sorts of flags in our own linguistic community. If someone begins a statement with "I sort of think…", or "I'm inclined to say…" we are more likely to categorize their speech as having opining force rather than assertive force. Importantly, notice that if someone says "my opinion is…" then often they will have made an *assertion* to the effect that they have a certain opinion.

as ubiquitous as the practice of transferring knowledge through testimony or sharing our opinions with one another. Other times it could be as narrow and specialized as a game of baseball that licenses declaring someone "out," or a religion than enables the act of granting "absolution." Certainly the practice of storytelling seems to fit well with this criterion, though it is not something that Gale spends much time discussing. The internal goals and norms of storytelling explain why we would need a fictive mode of speech.

Sometimes what guarantees the important role that speech acts are able to play within our practices is the way many of them enable us to communicate our internal states to one another. In particular, the fact that a speech act is constituted by certain norms means that speakers can use it to reliably communicate to one another that they have satisfied those norms. As we discussed in the previous section, this is how speakers who assert express their justified beliefs and speakers who promise express their intentions.

The reason that speakers can use these speech acts to reliably signal their inner states is that they undertake *risks* by so signaling.[126] If it is discovered that they do not satisfy the felicity conditions of the speech acts they are performing, they are subject to criticism. In the event that their flouting of these constitutive norms is systematic enough, they may face serious social consequences, including becoming unable to achieve the characteristic effects of the practice altogether. After enough false alarms, the boy who cried wolf was cut off from the ability to successfully transfer knowledge by performing the speech act of assertion—with dire consequences indeed.[127]

---

[126] For a development of this connection between risk taking, cost incurring, and action expression see Green (2007) chapter 3, especially section 5.

[127] A distinct but related phenomenon, *silencing*, involves structural inequalities that allow privileged speakers to prevent marginalized speakers from being full participants in certain social practices, partly be cutting them off from the unencumbered

How does the Hybrid Speech Act Theory's positing of hitherto unrecognized normative speech acts fare on these criteria? In our discussions of Copp and Finlay in the last chapter, we noted that the expression of belief and motivation by speakers who make normative claims should be analyzed as arising neither from the semantics of normative terms, nor from the general conversational principles and maxims. In the former case, this was because the expression does not survive in "force stripping" or "embedded" contexts.

In the latter case, this was because the expression is not explicitly cancellable in the Gricean sense. Not only can we use sentences containing normative terms to perform *non-normative* speech acts (as the examples of embedding show), we can also perform *normative* speech acts using sentences that do *not* contain the relevant normative terms. For example, I could successfully make a moral claim by uttering the sentence "there are no *thieves* among *my* friends" or a claim of etiquette by uttering "*I* place the fork on the *left* side." Normative claims exhibit independence of force and content, the first step towards speech act-hood.

To what extent are the constitutive norms governing these alleged speech acts distinct? This is the first criterion. In keeping with the broad insight of the hybrid approach to metanormative theory, HSA posits that normative speech acts require speakers to possess both cognitive and moving states. However, HSA also leaves open the question of

---

use of certain speech acts. Here the disadvantaged speaker has done nothing to disqualify themselves, but nevertheless is perceived as disqualified by the community. Because they do not have enough social capital to take an adequate risk, they are rendered incapable of accomplishing the typical perlocutionary effects of their speech acts. Importantly, they still count as having performed the speech act in question. It's just that other's don't take them to have done so. In this respect my understanding of speech acts parts ways with theories that require audience uptake—for example see Kukla (2014). My own reasons for emphasizing speaker intention rather than audience uptake will become clear in the next chapter. For more on silencing and epistemic injustice see Fricker (2009), Chapter 1.

just how many distinct normative speech acts there are. This will depend on the extent to which the different normative domains are designed for speakers to show different attitudes to one another.

For example, the moral and the aesthetic plausibly pick out different speech acts, given that moral claims require a speaker to have certain inclinations towards *actions* while aesthetic claims require speakers to have certain inclinations towards *feelings*. Is there a distinction between *ethical* speech acts and *moral* ones? Perhaps not, but the question requires more investigation. What all normative speech acts have in common is that they are designed for use by speakers who possess mental states of both "directions of fit," and what distinguishes normative speech act types from one another is the precise nature of these required psychologies.

The second criterion is more difficult to assess. It would be unusual for a speaker to begin a speech act with "I hereby *morally* judge…" or "I hereby claim *aesthetically*…" Nevertheless, it *is* intelligible for speakers making normative claims to qualify or clarify after the fact, especially when they use thin normative terms. "I only meant that is what Tom 'should' do in the *prudential sense*" or "The food is 'good' in that its source is ethically responsible, but it certainly isn't delicious" would be natural statements for a speaker to make when wishing to avoid misunderstanding. In other words, conventions do exist that allow us to explicitly flag which sort of normative speech act we are performing.

Finally, according to HSA each of the distinct normative domains of speech finds its home within a distinct practice (criterion three). Though these practices may overlap considerably (just take etiquette vs. the aesthetic, or the prudential vs. the rational), as long as they have somewhat different speech act rules and somewhat different aims they will

count as distinct. Epistemic norms aim us towards justification and truth, which is why being able to claim that a belief formation process is 'good' in the epistemic sense is an important part of epistemic practice. It commits a speaker to certain patterns of behavior—methods for weighing evidence and forming credences. Similarly, aesthetic norms aim us toward having an appropriate affective reaction to the experience of an object. The ability to show one another what we have felt by sharing our aesthetic judgments is an integral part of aesthetic practice. What the normative domains have in common with each other (and with the speech act of assertion) is that they involve a cognitive requirement in addition to the affective/motivational one. This is how HSA "has it both ways."

The hybrid speech act view's positing of hitherto unrecognized speech acts has good justification according to the criteria we have identified. Not only that, but it is a hybrid theory that can fully capture both the cognitive and practical elements of our normative speech. By paving the way between general conversational rules and the conventional meanings of terms, HSA identifies a source of expression that is neither explicitly cancellable nor survives in embedded contexts. However, there is much more to be said about the content of the different normative speech act felicity conditions. In the next two sections I look at the examples of moral speech acts and aesthetic speech acts in particular. I show how my hybrid speech act framework makes sense of some outstanding puzzles in the literature, and in doing so demonstrate the sorts of considerations that might bear on the project of filling out the felicity conditions of the different normative speech act types.

Section 3: Michael Smith's Moral Problem Paradox

In Chapter 1 we briefly discussed a long standing puzzle in metaethics. The most famous form of this puzzle is the one described by Michael Smith as the "moral problem paradox." The paradox consists in the fact that the following three claims, while independently plausible, appear to be mutually incompatible:

A. Moral judgments of the form 'It is right that I Φ' express a subject's beliefs about an objective matter of fact, a fact about what it is right for her to do.

B. If someone judges that it is right that she Φs then, *ceteris paribus*, she is motivated to Φ.

C. An agent is motivated to act in a certain way just in case she has an appropriate desire and a means-ends belief, where belief and desire are, in Hume's terms, distinct existences.[128]

Here Smith intends A to represent cognitivism, B to represent internalism, and C to represent Psychological Humeanism. In Chapter 1 I pointed out that these three claims are only inconsistent if one takes 'judgments' and 'judges' in A and B respectively to refer to beliefs and the act of believing. If moral judgments *are* beliefs, then it is true that one must decide between Psychological Humeanism and internalism. But there is no good reason to think that one must identify moral judgments with beliefs in order to count as a cognitivist.[129] The heart of the cognitivist's view is, as stated in A, that moral judgments *express* beliefs (and hence take truth assessable propositions as their objects). After

[128] Smith (1994), p. 12.

[129] For example, Russ Shafer-Landau simply stipulates that moral realism and moral cognitivism directly imply the claim that moral judgments are beliefs. See Shafer Landau (2003), Introduction and Chapter 1 (especially pp. 4, and 17–20).

working through the details of HSA in this chapter, we are now in a position to see that a fully satisfying hybrid treatment of the puzzle is possible.

The hybrid speech act view is committed to the following version of motivational judgment internalism: necessarily, if a speaker performs a moral speech act in all sincerity, then she is at least somewhat motivated to act in accordance with what she says. Moral claims, when made from within the moral domain of discourse, have the force of a moral speech act—a speech act whose constitutive norms require wholehearted speakers to instantiate both cognitive and moving states. At the same time, HSA may also do justice to the intuitions of motivational judgment externalists. Typically, these intuitions hinge on the possibility of the "amoralist"—a character who has a moral belief, sincerely asserts the moral claim associated with that belief, and yet lacks any appropriate motivation whatsoever. The reason the hybrid speech act view can accommodate these intuitions is that it is consistent with Humeanism.

Psychological Humeanism states that beliefs and desires are distinct types of mental states, the possession of neither of which necessarily implies the other (though we may often expect certain belief-desire pairs to come together). Thus in order for any metaethical view to be consistent with Humeanism it must admit that it is *possible* for a person to, for example, have the belief that eating meat is wrong yet possess no disinclination whatsoever towards eating meat. It is also possible for a speaker to *assert* that it is wrong to eat meat, have no negative attitudes towards eating meat, and yet count as fully sincere in what he says (as long as he believes it). I may even find myself in a situation in which my audience knows that I am merely asserting this moral proposition, and thus is not confused by my disavowals of motivation. For example, my audience may know that I am suffering from

severe depression, or from a personality disorder, or perhaps that I have just read some work of moral philosophy that argues deductively from premises I find compelling to a conclusion about which I am thoroughly apathetic.[130]

The hybrid speech act view diagnoses all of these sorts of situations as ones in which, for one reason or another, a speaker has failed to achieve full participation in the moral domain of discourse—moral topics are being discussed, but in an impoverished way. To converse outside the moral domain is to be engaged in a conversation made up of speech acts with constitutive norms such that speakers who perform them are not criticizable for lacking the motivational states typically associated with the moral predicates they are applying (e.g. asserting, opining, conjecturing, supposing for the sake of argument, etc.). Of course, there may be situations where such a conversation is all the interlocutors are capable of, or where it is sufficient for accomplishing the goals they have set themselves. However, such episodes still amount to falling short of the constitutive aims of distinctively *moral* discourse.

More generally, HSA is committed to there being a *cost* to remaining outside properly normative domains of discourse, if that is what one finds oneself doing. If a group begins systematically using, say, epistemic or aesthetic normative terms in conversations intended *merely* to express their beliefs to one another (or perhaps *merely* to trace the logical relationships between concepts) then they have cast aside one of the central resources that normativity offers. In fact, my view implies that such a practice has ceased

[130] By this last case I envision something along the lines of Raimond Gaita's discussion of Peter Singer. Often such philosophical discussions can "lack moral seriousness." See Gaita (2004) pp. 55–61.

to be normative altogether, at least in any substantive sense.[131] Whatever other goals such a discourse might accomplish, it is limited in its ability to contribute to living well.

It is a sociological question whether a great deal of normative ethics is taking place outside the moral domain of discourse. That is, whether it has become widely accepted in contemporary analytic moral philosophy that when we talk about first order ethics we may be considered fully serious merely by believing what we say, regardless of our deeper motivational commitments towards certain patterns of behavior. If so, moral philosophy may come to have an effect similar to that of Kongzi's student Zai Wo:

> Zai Wo was sleeping during the daytime. The Master said, "Rotten wood cannot be carved, and a wall of dung cannot be plastered. As for Zai Wo, what would be the use of reprimanding him?"
>
> The Master added, "At first, when evaluating people, I would listen to their words and then simply trust that the corresponding conduct would follow. Now when I evaluate people I listen to their words but then closely observe their conduct. It is my experience with Zai Wo that has brought about this change."[132]

To the extent that we consider such a situation to be an unfortunate one, and recognize it to be our own, the hybrid speech act view can show us where we have gone wrong. If we are in fact headed down such a path, then this may be one of those odd places where Confucius, David Hume, and Raimond Gaita might find themselves equally concerned about our 21st century ways of speaking, and for similar reasons.[133]

---

[131] I have in mind here something like the considerations raised in Gibbard (1990), p. 33.

[132] Confucius, *Analects* 5.10 (Slingerland 2005, p. 14).

[133] See especially *Analects* (1.4, 4.22, 9.24, 13.3, 14.27, and 18.8), *Treatise* (3.1.1.5) and *Good and Evil* (pp. 321–322).

Section 4: The Puzzle of Aesthetic Judgment[134]

Recent literature on the norms of aesthetic assertion has uncovered a puzzle that goes straight to the heart of our understanding of aesthetic knowledge and the purpose of aesthetic discourse. Simply put, this puzzle consists in three independently plausible observations that, taken together, are mutually inconsistent.[135] They are:

1. If a speaker performs an unqualified assertion concerning the aesthetic properties of an object, but has had no first-hand experience of the object itself, her act is judged to be *improper*.[136]

2. It is possible to come to know that an object has particular aesthetic properties on the basis of testimony alone.

3. It is sufficient for performing an assertion properly that one know the proposition one is asserting.

The hybrid speech act view suggests a new solution to the puzzle, one that turns on recognizing an ambiguity in statement 1. Inconsistency in the triad only arises from taking aesthetic "assertions" to be the same type of speech act as ordinary assertions. As we have just seen, the hybrid speech act view analyzes "aesthetic assertions" as actually being a previously unrecognized and distinct species of speech act from ordinary assertions, differentiated by a unique set of linguistic-pragmatic norms. These norms are explained by

---

[134] Portions of this section first published in Morgan (2017).

[135] Most recently Robson (2015).

[136] For the purposes of this discussion I am restricting myself to what are known as "verdictive" aesthetic properties. These are normative properties like beauty, ugliness, aesthetic merit, aesthetic defectiveness, etc. I do not include so called "substantive" aesthetic properties like daintiness, dumpiness, elegance, gracefulness, delicacy, balance, warmth, awkwardness, sadness, etc. except insofar as these substantive properties have come to represent *thick* normative aesthetic properties in particular corners of aesthetic practice. If a substantive aesthetic property is commonly understood to be non-normative, then it is not included in this discussion.

the constitutive function of aesthetic discourse to both transfer information and express affective states.

Before we jump to the details of the HSA solution, let's take a moment to consider why we find these three theses so plausible. Consider the following story:

> As we sit down to dinner, we happen to be discussing our favorite films and film directors. I bring up Japanese film, praising Yasujiro Ozu for his ability to elevate mundane events and ordinary objects. You counter that Kenji Mizoguchi's poignant portrayals of humans suffering across both historically informed backdrops and in fantastical situations should put him higher on our list of best directors. Attempting to secure your position, you emphatically state:
>
> "Just take *Ugetsu*—it's a classic ghost story in the *jidaigeki* genre. Besides, it's an extraordinarily beautiful piece of filmmaking."
>
> Having never seen *Ugetsu* myself, and with very limited experience of Mizoguchi in general, I trust your judgment and concede the point; the conversation moves on to other topics. Later that evening, however, our mutual friend takes me aside and informs me that you have never seen *Ugetsu* either. It's not a big deal, but I do feel misled, so I ask you about it. You readily admit that you have never seen the film, not a single scene.
>
> "Fine," I respond "I appreciate your honesty, and I don't even really mind your raising it as an example—I'm sure you're right. But what did you mean when you said it was 'a beautiful piece of filmmaking'?"
>
> "Oh that," you say. "Sure, I haven't seen the film, but I *know* that it is beautiful. I read about it in a book on Mizoguchi's work. It is widely regarded as the definitive analysis, and so far has been spot on with every film I've seen. In fact, *Ugetsu* is the only one I haven't gotten to yet."
>
> I value our friendship, and your explanation goes some way towards assuaging my confusion, so I let the matter drop. Nevertheless, I am left with the lingering sense that in making your initial statement you misrepresented yourself. What you should have said is that you've *read* that it is a beautiful piece of filmmaking.

I take this vignette to show that it is infelicitous for a speaker to claim that some object has an aesthetic property, in a situation where the speaker has had no first-hand experience of that object. "Infelicity" is meant in the same way we have been using it throughout this chapter, to single out J.L. Austin's distinctively linguistic-pragmatic form of impropriety.

Of course, our aesthetic claims could be improper for many other reasons in a particular situation. They may be impolite, immoral, imprudent, etc. We may even find ourselves in situations where another norm *recommending* that we state our claim can outweigh the linguistic-pragmatic considerations against it. However, given only the goal of successful communication, there are norms that govern our speech acts. These are the felicity conditions we have been speaking about throughout this chapter.

Determining the precise set of linguistic-pragmatic norms that govern the speech act of assertion is still an open and much debated issue. Nevertheless, it is fairly plausible that *knowledge* of what one asserts is *sufficient* for counting as having asserted in good faith. This is claim (3) in our triad. Perhaps all that is required is that one's belief in the statement be justified, or that one have a reasonable belief *that* one knows. Whichever way that debate is settled, what the vignette above shows is that even when the strongest of the candidate norms of assertion are satisfied (having the true belief that one knows), an *aesthetic* claim may still be judged out of order.

Of course, there is one more claim required to generate this puzzle. That is (2) above: that it is possible to come to know some of the aesthetic properties of an object on the basis of testimony alone. While this claim is plausible, our confidence in it may ultimately depend on the prospects for solving the puzzle. Given no other route to holding on to (1) and (3), many might be willing to give up on (2). I will return to the most popular example of this strategy, the neo-Kantian view, below. First let us turn to the details of the HSA solution.

According to the hybrid speech act view, aesthetic judgments and ordinary assertions are two distinct types of speech acts, precisely because they are governed by

different norms. Knowledge obtained via testimony can license speakers to make assertions about matters that they have not directly observed. In contrast, it is appropriate to criticize speakers making aesthetic judgments for ascribing (without qualification) aesthetic properties to objects that they have never experienced, even if the speaker knows the aesthetic claim is true on the basis of testimony.[137] This is explained by the fact that aesthetic discourse is a social practice with certain goals and possibilities. Aesthetic judgment is a speech act governed by constitutive norms distinct from the norms governing ordinary assertion, norms that allow it to play an important role within our aesthetic practices.

Just like everyday assertion, aesthetic judgment is governed by an epistemic norm that enables aesthetic judges to transfer knowledge (or at least justified belief) to one another via testimony. However, in contrast with everyday assertion, aesthetic judgment is governed by an additional norm. This norm says that, whatever aesthetic property one is predicating of the object, one's aesthetic assertion is appropriate *only if* one has had an affective reaction implied by that predicate in response to an experience of that object. This norm allows aesthetic assertion to play the important functional role within the practice of aesthetic discourse of enabling us to *express* our affective states.

Notice that HSA's solution is stronger than what is needed to solve the original puzzle. Not only does it claim that aesthetic asserters are required to have experienced the

[137] I would like to leave it open exactly what it takes to experience an aesthetic object. There is certainly some vagueness here, but there are also clear cases. Perhaps someone who has seen high quality photographs of the *Mona Lisa* shouldn't be criticized for saying that it is beautiful, while someone who claims that the acting in *The Room* was awful without ever having watched a single scene the film should be—even if they both say only what they know to be true.

objects they are speaking about, but also that they must have had an affective reaction associated with the properties they are predicating of those objects. This is in accordance with HSA's status as a fully *hybrid* theory. But we should not pursue hybridity for its own sake. To see why it is justified in the realm of aesthetic discourse, compare the following sentences:

(A) *The puzzle case*: "The alps are truly beautiful. Of course, I have never seen them—not even a picture. Nevertheless, I say they are beautiful."

(B) *Flouting belief condition*: "The alps are truly beautiful. Of course, I don't believe that they are beautiful. Nevertheless, I say they are beautiful."

(C) *Flouting affect condition*: "The alps are truly beautiful. Of course, I really felt nothing whatsoever when I saw them. Nevertheless, I say they are beautiful."

All three of these statements exhibit a similar sort of absurdity. In the absence of a further explanation, speakers of such sentences will leave their audiences with lingering puzzlement and perhaps even outright confusion. This indicates that in addition to the first-hand experience condition, the belief and affect conditions give rise to a form of expression that is *noncancellable*. Thus it is not analyzable as arising from conversational implicature. At the same time, the expression of belief, affect, and first-hand experience does not survive embedding ("Jerry thinks the alps are beautiful", "if all the alps are beautiful, then the Grossglockner is beautiful") and thus cannot be a form of conventional implicature.

Notice that I have not made explicit what "feel" amounts to in this context. It seems plausible to me that there are a variety of affective reactions a speaker could have had that an audience would count as sufficient for her beauty claim to be felicitous. What would be insufficient would be a total absence of affective response whatsoever. There are clear examples of *inappropriate* affective responses as well. Consider a speaker who says, "the

Alps are beautiful, though I felt disgust when I saw them." In the absence of an adequate explanation, such a claim will confuse the audience. Of course, adequate explanations are possible: "I smelled a rank odor at that exact moment" or "the sight brought up a disturbing memory."

According to HSA, these phenomena are best explained by the constitutive speech act norms of the various speech acts that belong to the different normative domains—and aesthetic judgment is no exception. What is interesting about the example of aesthetic assertion as opposed to, say, ethical assertion, is that "good faith" (full felicity) requires aesthetic asserters to have formed their affective states in response to the experience of a particular object. In contrast, good faith ethical asserters need only be motivated to act in accordance with their judgments—the causal story of how that motivation arose is (perhaps) less important.

The strengths of HSA's treatment of aesthetic judgment may be highlighted by contrasting it with two alternatives in the literature. The first is John Robson's Creative Signalling account.[138] Robson proposes to solve the puzzle by positing that one of the constitutive functions of aesthetic discourse is for us to show each other our positive traits. In a nutshell, his account depends on the view that our practice of aesthetic creation originated as an evolutionary adaption—specifically as the result of sexual selection. In the distant past, our creating of attractive artworks thereby demonstrated that we ourselves possessed the attractive fitness enhancing characteristics (dexterity, attention to detail, etc.) required to produce such works. Robson then claims that aesthetic *appreciation* requires the possession of traits similar to those required for aesthetic *creation*. Hence, one function

[138] Robson (2015).

of aesthetic discourse is for us to make manifest these traits. The practice of aesthetic assertion advances this function only if it licenses us to criticize speakers who have not experienced the objects about which they speak. Such "bad faith" speakers represent themselves as having traits they have not manifested—at least not with respect to the object in question.

Robson readily admits that his solution will only be attractive to the small subset of people who are antecedently disposed to accept the evolutionary account of our practice of aesthetic creation. But even setting the controversial nature of this theory aside, Robson's view encounters further difficulties that my own approach overcomes. For one thing, Robson never discusses whether he takes aesthetic assertion to be a distinct speech act from everyday assertion. This makes it unclear in exactly what way he takes the three claims of the puzzle to be intuitively inconsistent, and leaves the details of his solution vague. In addition to these shortcomings, Robson has at best uncovered the impropriety at work in a very small subset of aesthetic discourse. That is to say, while showing certain aesthetic traits may be a goal of conversations between professional and self-styled aesthetic critics, it hardly seems to the purpose of ordinary aesthetic conversations. For example, consider:

> (D) *Flouting Aesthetic Authority*: "Van Gogh's *Starry Night* is truly beautiful. Mind you, I do not claim to have any refined skill of aesthetic appreciation. Nevertheless, I say that it is beautiful."

Such a statement might generate some absurdity coming from the mouth of a critic in the context of the professional art establishment. I am entirely open to the idea that in some contexts additional pragmatic norms are at play. In everyday aesthetic conversation

however, the speaker in (D) succeeds perfectly well at conveying to her audience both the information (she believes that the painting is beautiful) and the affection (she has felt awe in response to experiencing the painting) that are distinctive of beauty claims. Were a friend of mine to say such a thing upon returning from a trip to the Museum of Modern Art in New York, I wouldn't feel puzzled in the slightest. Robson's theory has no explanation to offer of these everyday cases.

The second sort of view I will briefly address is that of neo-Kantian orthodoxy.[139] Views along this line can take many forms, but what they all have in common is their commitment to the thesis of Pessimism.

> *Pessimism*: One cannot achieve knowledge (or perhaps proper belief) concerning aesthetic matters on the basis of testimony alone.[140]

Pessimism is, simply put, just the denial of (2). One motivation for such a view, as the name suggests, is a lack of optimism about alternatives for solving the puzzle: given that we have greater confidence in statements (1) and (3), if we must give up one claim it will have to be (2). I take myself to have undermined this motivation by providing a plausible interpretation under which the three claims are consistent, and an explanation for why this interpretation is correct. The other motivation for Pessimism consists in independent Kantian arguments for a particularly stringent view of aesthetic knowledge. For example, such a view might be committed to the idea that aesthetic knowledge of an object requires a certain kind of acquaintance with that object. Fortunately, I may remain entirely neutral about such views, as the solution I advocate here is consistent with any of them.

---

[139] For just one example see Hopkins (2011).
[140] Taken from Robson (2015), p. 2.

Importantly though, if such a view is correct it must go on to provide an explanation of why we intuitively *think* that (2) is true. That is to say, the neo-Kantians owe us an error theory for why we find the puzzle of aesthetic assertion so puzzling.

To sum up then, if the neo-Kantian is wrong we need a solution to the puzzle of aesthetic assertion. HSA offers a highly plausible "optimistic" solution of the puzzle, which is grounded in a broader hybrid metanormative theory. If on the other hand the neo-Kantian is right, her view is controversial enough that we shouldn't expect to read it off of the linguistic-pragmatic norms of our everyday aesthetic conversations. A solution to the puzzle would still be needed—why do (1)–(3) independently *seem* so plausible? The answer is that aesthetic "assertions" aren't really assertions after all. Because they are governed by different felicity conditions, they constitute a distinct type of speech act. This speech act allows speakers to share with one another both their beliefs, and the affective reactions they have experienced in response to particular aesthetic objects.

Part IV: A Coda for Bar-On and Boisvert

Section 1: Bar-On's Ethical Neo-Expressivism

In a series of co-authored papers, Dorit Bar-On has applied her neo-expressivist view of avowals described above to the case of ethical claims. Because her view is similar to mine in significant ways, it is important to highlight the differences between our theories. I will also show why friends of the hybrid approach should prefer my own hybrid speech act view. This section begins by giving a brief overview of Bar-On's account of ethical

claims, and then moves to discussing the relationship between her view and hybrid approaches more generally and my own view in particular.

In their 2009 paper "Ethical Neo-Expressivism," Bar-On and Chrisman begin by applying Bar-On's distinction between what a speaker action expresses$_1$ by performing an utterance, and what is semantically expressed$_3$ by the vehicle of the utterance (the sentence). In the case of a speaker who makes an ethical claim, this means distinguishing between the speaker's action of uttering the sentence with ethical content, and the ethical sentence itself. According to Bar-On, classical ethical expressivists like Ayer were right to say that one of the primary purposes of ethical discourse is to *express* motivational states. However, ethical expressivism went wrong in conflating the *action* and *semantic* varieties of expression.

Ethical Neo-Expressivism (ENE hereafter) holds on to the idea that speakers who utter ethical claims express their motivational states, but clarifies that this expression is of the *action* variety. Sentences with ethical content, on this view, may still semantically express$_3$ truth apt propositions. Thus we may get all the virtues of ethical expressivism (a strong connection between ethical discourse and motivation) with none of its vices (Frege-Geach problem, clash with intuitions, failure to explain disagreement, lack of appreciation for declarative surface grammar, etc.). This is accomplished by clearly separating what speakers express$_1$ *through their acts* from what the *vehicles* of these acts express$_3$. So far, so much like my own view.

According to Bar-On and Chrisman, what is the mechanism by which this action expression$_1$ is produced? The answer to this question is less clear, given that they never specifically describe the act of making an ethical claim as a *speech act*. They refrain from

making this commitment because they wish their view to be neutral between the action of making an ethical claim in thought versus the action of making an ethical claim in speech.[141] However, they do say that these "ethical claim acts," whether they occur in speech or in thought, require in their "felicity" or "propriety" conditions that the speaker (or thinker) of the ethical claim be in the corresponding motivational state.[142] This suggests that they would be friendly to the more developed account I pursue.[143]

Throughout this chapter I have been working to clarify what it means for a speaker to express$_1$ a mental state by performing a speech act that has the possession of that mental state as one of its felicity conditions. The clearest and most complete account comes from Green's work on non-conventional non-conversational implicature. Reading charitably, it is to be expected that Bar-On and Chrisman may avail themselves of this account as well, at least for the ethical claim acts that speakers perform in speech. But what about the ethical claim acts that thinkers perform in thought? How may thoughts be said to have "felicity conditions," much less involve action expression$_1$ by the thinkers thinking them in virtue of those felicity conditions?

This is an important question, and I believe one of the most difficult issues facing hybrid metanormative theories. It is the goal of the next chapter to show that my own hybrid speech act view can answer it. I believe that the solution lies in clearly articulating the way that the mechanism of expression behaves in speech, and arguing that surrogates of this mechanism are present in some modes of thought as well. But developing such an argument is a substantive, controversial, and delicately performed task.

---

[141] Bar-On and Chrisman (2009), p. 141.
[142] Bar-On and Chrisman (2009), pp. 141, 143, and 146.
[143] A suspicion confirmed in conversations between myself and Dorit Bar-On.

Bar-On and Chrisman, by keeping their discussion of "ethical claims" at a level of generality that does not distinguish speech act from thought act, are unable to make this move. Hence they leave both phenomena under-explained. Thus while my hybrid speech act view is broadly consistent with ENE, it goes much further—both in depth and in scope. The hybrid speech act view goes deeper by explaining the mechanism by which expression$_1$ takes place both in speech (non-conventional non-conversational implicature) and in thought (in the next chapter). The hybrid speech act view also reaches further in scope by including all normative domains, not just the ethical. As its name implies, Ethical Neo-Expressivism is only a theory about discourse and thought in the ethical domain. While it certainly could be extended in principle, as yet its proponents have not done so.

As far as it goes, ENE is similar to the hybrid speech act view, with one important caveat: ENE takes speakers who make ethical claims to be expressing$_1$ *only* their motivational states, while my hybrid speech act view takes normative speakers to be expressing$_1$ both motivational states and beliefs.[144] For this reason, as Bar-On et al. fully admit, ENE is not a truly *hybrid* metaethical theory (much less a hybrid *metanormative* theory). It is a form of moral judgment internalism that captures key insights of ethical expressivism, without the serious drawbacks of ethical non-factualism. And it has much in common with the hybrid speech act view. However, given that it is not a hybrid theory, it is not a true competitor.

Why don't Bar-On et al. analyze ethical claims as requiring—thereby enabling speakers to express$_1$—beliefs, in addition to motivation? This question opens the way for an important clarification. In their 2014 essay "(How) Is Ethical Neo-Expressivism a

[144] Bar-On and Chrisman (2009), p. 145.

Hybrid View?" Bar-On et al. lament the proliferation of the so-called "ideationist" conception of meaning in metaethics, and the turn it has taken specifically in many hybrid theories:

> That is, they suggest that *all* sentences mean what they do in virtue of the "idea" (mental state type) they express; is it just that ethical sentences express a different kind of mental state from descriptive sentences. Recently, a further epicycle of this debate has seen the articulation of various "hybrid" views that in some way seek to get the best of both views that ethical sentences express conative attitudes and that ethical sentences express beliefs by giving the meaning of ethical sentences in terms of both cognitive and conative states.[145]

For example,

> According to Daniel Boisvert's "Expressive-Assertivism"…the claim "Tormenting the cat is bad" expresses both the belief that tormenting the cat has a certain (non-speaker-relative) property and a negative attitude toward things with that property. The claim therefore has the following meaning, according to Boisvert (2008, 172): "Tormenting the cat is F; boo for things that are F!"[146]

Bar-On et al. point out that ideationist theories of meaning *in general* have been rejected by philosophers of language on account of "insurmountable difficulties."[147] Part of the reason is that *sentences* just do not seem to be in the business of "expressing" (in any sense of the term) mental states. Sentences may semantically express$_3$ propositions that *refer* to mental states, and speakers may use sentences to perform *actions* that action express$_1$ mental states—but sentences themselves do not directly "express" mental states.

---

[145] Bar-On et al. (2014), p. 224.
[146] Bar-On et al. (2014), p. 224.
[147] Bar-On et al. (2014), p. 224.

Perhaps part of the reason Bar On et al. do not go in for the fully hybrid view is that they wish to distance themselves from a dead-end ideationist strategy. However, I argue that in doing so they have thrown the baby out with the bathwater.[148] As discussed in chapter 1, we have all the reason in the world to want a theory of normative speech that analyzes speakers as expressing *both* moving and cognitive states. And, as noted above, Bar-On herself observes that there is "no general reason to maintain that one cannot express$_1$ more than one state with a bit of behavior."[149] The hybrid speech act view avoids the pitfalls of ideationist semantics, while embracing a fully hybrid understand of what speakers action express$_1$ in making normative claims.

Section 2: Boisvert's Expressive-Assertivism

I'd like to close this section by briefly clarifying Dan Boisvert's view: his motivations, to what extent his theory counts as an instance of the ideationist approach, and why I take my hybrid speech act view to be preferable to his "Expressive-Assertivism." In some ways, the two views are very similar—specifically, both analyze normative speakers as simultaneously expressing practical and cognitive states. In this sense Boisvert's view is a truly hybrid theory (in contrast with Bar-On et al.'s ENE) and thus more of a true competitor. However, there are significant differences between our views as well.

---

[148] In fact, it may be that they have confused the baby for the bathwater. In Bar-On et al. (2014) they seem to claim that a commitment to ideationist semantics is *essential* to hybrid views (see especially pp. 243–244). Regardless of whether most or all hybrid theorists were originally motivated by a commitment to ideationism, what the hybrid speech act view shows is that there *is* space for a meaningfully *hybrid* view that is semantically *neutral*.

[149] Bar-On (2004), pp. 306–307.

The clearest and most recent presentation of Boisvert's view comes in his 2014 essay "Expressivism, Nondeclaratives, and Success-Conditional Semantics." In that essay his goal is to explore the prospects for metaethical expressivists to exploit the resources of success-conditional semantics (SCS) in avoiding the Frege-Geach problem. What unites expressivist views is that they all analyze the connection between moral language and the expression of practical states in terms of semantics. Traditionally, expressivists have incorporated this into a more general ideationist (or "mentalist") semantics, according to which the meaning of a sentence is just the mental state it expresses. As noted in the previous section (and briefly in the previous chapter), I follow Green and Bar-On in viewing such theories as being at best confused.

Importantly, Bar-On et al.'s comments quoted above notwithstanding, Boisvert himself *agrees* that the ideationist strategy is a dead end—hence his exploration of the prospects for expressivists to make use of success-conditional semantics instead. Essentially, SCS is the view that the meaning of a sentence is given by the success conditions of the speech act that the sentence would typically be used to perform. According to Boisvert, each sentential mood (declarative, imperative, interrogative, etc.) corresponds to a speech act family that it is typically used to perform. In turn, each family of speech acts has a typical conversation purpose that determines that speech act's criteria for "success." Here's an example:

> Declarative sentences are apt for performing direct assertives. SCS therefore aims to construct a theory of success for a language from which one can derive theorems specifying its declaratives' truth conditions, which they inherit from the direct assertives they are apt for performing. Taking the declarative "The street lights are on" as an example and relativizing to a speaker and time of utterance, which we will abbreviate using the

> subscript "$_{[s,t]}$," an adequate success-conditional meaning theory for English would permit the following abridged derivation:
>
> (i) "The street lights are on" is successful[s,t] in English iff "The street lights are on" is true$_{[s,t]}$ in English;
>
> (ii) "The street lights are on" is true$_{[s,t]}$ in English iff the street lights are on at the time of utterance; (therefore)
>
> (iii) "The street lights are on" is successful$_{[s,t]}$ in English iff the street lights are on at the time of utterance.[150]

Boisvert carefully goes through this derivation exercise for each of the sentential moods and its corresponding speech act. For our purposes here the only other moods that will be important are those of the exclamatives and optatives, both of which Boisvert claims are "apt for performing" speech acts from the family of expressives. So the semantic meaning of an exclamative sentence like "hooray!" would be the excitement of the speaker, because that is the state of mind demanded by the success condition of the expressive speech act that "hooray!" is typically used to perform.

It is essential to Boisvert's SCS that the sentential moods are tied to the speech act families in this way. The obvious problem is, of course, that we can use the same sentence to perform speech acts from different families. A sentence in the declarative mood may be used to ask a question, a sentence in the interrogative mood may be used to give a command, an exclamative may be used to make an assertion, etc. Here is how Green puts the problem:

> [G]rammatical mood together with content underdetermine force. 'You'll be more punctual in the future' is in the indicative grammatical mood, but as we have seen, that fact does not determine its force. The same may be said of other grammatical moods. Although I overhear you utter the words,

[150] Boisvert (2014), p. 32.

> 'shut the door', I cannot infer yet that you are issuing a command. Perhaps instead you are simply describing your own intention, in the course of saying, "I intend to shut the door." If so, you've used the imperative mood without issuing a command. So too with the interrogative mood: I overhear your words, 'who is on the phone.' Thus far I don't know whether you've asked a question, since you may have so spoken in the course of stating, "John wonders who is on the phone." Might either or both of initial capitalization or final punctuation settle the issue? Apparently not: What puzzles Meredith is the following question: Who is on the phone?[151]

How does Boisvert explain this phenomenon? He appeals to the distinction between performing a speech act directly vs. performing it indirectly. For example, when I use the declarative sentence "the door is shut" to ask a question (probably uttered with a rising intonation) Boisvert would describe me as having directly asserted that the door is shut, and by so doing indirectly performing the speech act of asking a question.

It is easy to see the motivation for the indirect/direct speech act performance distinction, given the goal of advancing an SCS style semantic theory. However, in the absence of SCS the distinction is merely an unmotivated epicycle in speech act theory, and a counterintuitive one at that. If I use a declarative sentence to ask a question, and in so doing perform an assertion along the way, why do I not express a belief that what I say is true? And why does my audience not criticize me for lacking this belief? This problem is compounded once it is recognized that a sentence may be used to *speaker mean* something quite different from its literal meaning, even when it is used to perform a speech act that does "correspond" (in Boisvert's sense) to its mood. For example, I may sarcastically utter the exclamative "hooray!" to perform the expressive speech act of venting my frustration and disgust at the occurrence of an event that I strongly resent. Is the meaning of my

[151] Green (2015).

sentence still excitement? Do I express both excitement and disgust? It is hard to see how the SCS can account for all the colorful uses to which we put our language.

Boisvert may have a response to this challenge. In turn there are more objections that could be raised.[152] However, the principle point I want to make is just that there is a theoretical cost to taking on board the direct/indirect distinction, just as there is to maintaining that each sentential mood corresponds to a particular speech act family. Of course, as Green says, "it is a plausible hypothesis that grammatical mood is one of the devices we use, together with contextual clues, intonation and the like to indicate the force with which we are expressing a content."[153] However, at the end of the day grammatical mood and semantic content together still underdetermine force. Both the direct/indirect distinction and the sentential mood/speech act family thesis are only motivated by an antecedent commitment to SCS. In turn Boisvert motivates SCS by arguing that, though it is ultimately unable to save classical expressivist theories, it does allow for one to salvage a hybrid form of expressivism.[154] This is the strategy he pursues in his theory of "expressive-assertivism" that was outlined briefly by Bar-On et al. in their 2014 piece. If Boisvert's expressivism is our only hope for capturing the practicality of normative thought and talk, then he thinks it is worth the theoretical costs SCS makes us pay. Given the

---

[152] For example, in both Boisvert (2014) and the original description of the view in Boisvert (2008), it is essential that there be a strong connection between the behavior of moral language and the behavior of slurs and pejoratives. It is this very connection that I cast doubt on in the last chapter. For additional problems see Sander (2016).

[153] Green (2015).

[154] See Boisvert (2014), pp. 38–42. I believe that Boisvert's discussion here is a good one. I would also simply like to point out that my own view (as well as Boisvert's) has the added advantage of being able to make sense of robust moral disagreement. This is in contrast to more traditional expressivist views, of even a more recent stripe like that of Michael Ridge (2014). For a more focused statement of the objection to Ridge, see Eriksson (2015).

success of my own hybrid speech act view, we are now in a position to see that making such concessions in not necessary.

I will close this discussion by summarizing the main points of disagreement between Boisvert and myself. According Boisvert's expressive-assertivism, a speaker who utters a sentence with moral content typically performs *two* speech acts directly. The speaker *both* performs an assertive *and* an expressive. Hence the meaning of a sentence with moral content is that the world is the way the sentence describes it as being *plus* the speaker's appropriate motivational attitude. When a speaker utters a moral sentence, meaning by so uttering to make a fully moral claim, they express their belief as well (in virtue of having performed an assertion). Boisvert clarifies that this makes his view a "hybrid" one along the following two dimensions:

> (H1) The direct illocutionary acts performed in correctly and literally using an ethical sentence are at least two: at least one assertive and at least one nonassertive.
> (H2) The meaning of an ethical sentence is both descriptive (truth conditional) and expressive; that is, its meaning is not exhausted by its truth conditions.

(H1) means that expressive-assertivism is what Boisvert calls "illocutionarily-hybrid," while (H2) commits it to being semantically-hybrid. To these two forms of hybridity, Boisvert adds a third:

> (H3) Moral thoughts are constituted by both a desire-like attitude and a representation-like belief; that is, they are logically complex.[155]

This thesis Boisvert describes as being "psychologically-hybrid."

---

[155] Boisvert (2014), p. 42.

To be clear, I deny the first two claims (H1) and (H2). I do not believe that moral sentences have a semantically encoded "correct and literal" usage (an implication of H1) any more than any other sentence does. It is no more correct to use a declarative moral sentence to ask a question than it is to use it to perform an assertion. This is connected to my rejection of success-conditional semantics and the accompanying direct/indirect speech act distinction (implied by H2). Finally, I am somewhat uncomfortable with the wording of (H3). I am committed to the existence of both moral beliefs and moral attitudes, and I do think that they often come together. However, I would not describe a person who happens to be in the conjunction of those two states as possessing some third "thought" that is composed by them. I don't believe that Boisvert wants to say this either.[156] But then the question arises what exactly creates pressure on normative thinkers to try to get themselves into both those states at once, aside from the novelty of counting as an example of (H3). I believe that explaining what is going on in conscious moral thought is a complex and difficult (largely empirical) question, which is why it is to precisely that issue that I will turn in the next chapter.

Up to this point our discussion has been, for the most part, limited to normative public speech. However, in order to fully capture what we are up to when we make use of norms, hybrid theories must be able to account for the cognitive and practical elements of normative thought as well. One way of putting this challenge comes in the form of Mark Schroeder's recent objection that hybrid views must abandon hope for consistency with broadly Humean theories of mind in order to maintain a form of motivational judgment internalism in the realm of normative thought. That is, he takes motivational judgment

[156] As was made clear by his questions during this dissertation's examination.

internalism to include a commitment to the view that "people who genuinely believe moral claims will be motivated."[157] We have discussed how the hybrid speech act view remains consistent with Humeanism, by admitting that it is possible for individuals to genuinely believe moral claims without being motivated. Does this mean that our commitment to internalism can run no deeper than public moral conversations? That would be an unsatisfactory conclusion indeed!

This challenge has, as of yet, been acknowledged only in a halfhearted way. Finlay's[158] and Copp's[159] comments on the subject, while suggestive, are hardly satisfying. The hybrid speech act theory can do better by connecting the metanormative discussion up with phenomena currently being studied by cognitive scientists and developmental psychologists. Demonstrating how will be the purpose of the next chapter.

[157] Schroeder (2014), p. 283.
[158] Finlay (2014), pp. 120–121.
[159] Copp (2014), p. 58.

CHAPTER 4

A TRULY HYBRID THEORY OF NORMATIVE THOUGHT

Part I: Introduction

> It looks to me as if, when the mind is thinking, it's simply carrying on a discussion, asking itself questions and answering them, and making assertions and denials. And when it has come to a decision, either slowly or in a sudden rush, and it's no longer divided, but says one single thing, we call that its judgment. So what I call 'judging' is speaking, and what I call 'judgment' is speech; but speech spoken, not aloud to someone else, but silently to oneself. What do you think?[160]

> We can never survey our own sentiments and motives, we can never form any judgment concerning them; unless we remove ourselves, as it were, from our own natural station, and endeavour to view them as at a certain distance from us. But we can do this in no other way than by endeavouring to view them with the eyes of other people or as other people are likely to view them….Bring [someone] into society, and he is immediately provided with the mirror which he wanted before.[161]

In this chapter I explain how the hybrid speech act theory's analysis of normative speech may be translated into an analysis of certain aspects of normative thought. In the

---

[160] Plato, *Theaetetus* 189e–190a (McDowell 1973, p. 75). For more on Plato's notion of judgment, see Burnyeat (1990). For connections with modern conceptions of inner speech, see Wiley (2006a) and Blachowicz (1997).

[161] Smith, *Theory of Moral Sentiments* III.1.3 (Raphael and Macfie 1982, p. 110).

next part I begin by describing inner speech, its phenomenology, distinctive character, and relationship to social speech. I also highlight the differences between the approaches of the two founders of modern research on inner speech, Jean Piaget and Lev Vygotsky. In Part III I present a survey of recent work on inner speech, with the goal of demonstrating the importance of a distinction between dialogic inner speech (including inner speech acts) and non-dialogic subvocal articulation (mere inner acts of speech). I emphasize the influence and continued importance of the Vygotskian approach to current investigations of inner speech, including my own.

In Part IV I argue that the theory of speech acts with which I have been operating thus far, Force Conventionalism (FC), is consistent with the possibility of inner speech acts. Because we have evidence for the existence of a dialogic mode of inner speech, and Force Conventionalism is the only theory consistent with such a possibility, this also consists in something of an indirect argument in favor of FC as the most adequate theory of speech acts. At the same time, because FC is already an independently plausible theory of speech acts in general, the fact that the possibility of *inner* speech acts is consistent with such a theory serves to legitimize the evidence canvased in favor of a dialogic mode of inner speech. Throughout this discussion I point out places where Vygotskian insights facilitate the integration of inner speech acts into the general theory of Force Conventionalism. The chapter concludes in Part V with an outline of the Hybrid Speech Act Theory's treatment of normative inner speech acts in particular, framed as a response to Schroeder's challenge raised at the end of the previous chapter.

In the interest of truth in advertising, I would like to begin by clearly describing what I *won't* be attempting to give a theory of in this chapter. As stated in the first chapter,

one of my overarching goals in this project has been to set forward the best hybrid theory that is consistent a general Humeanism about psychology. Recall that the two claims that I am taking to be essential to the Humean view are: (1) cognitive states and motivationally efficacious states are distinct from one another and (2) no cognitive state is conjoined with a motivational state *of necessity*. Remaining consistent with these two claims means that, in the present chapter, I will not be trying to show that normative beliefs are in any sense "hybrid."

Certainly I am committed to the existence of truly normative beliefs, and my standards for what it takes to count as having a normative belief are not very high. All one needs is a belief that includes as part of its content the predication of a normative property. In order to maintain consistency with Humeanism, I will assume for the sake of this chapter that such normative beliefs are *inert*.[162] That is to say, normative beliefs alone (unaccompanied by an appropriate practical state) are unable to move the believer to action *and* normative beliefs are not necessarily connected with any such motivational states. Of course, I take my view to also be consistent with the contrary: that normative beliefs *are* able to motivate without the presence of an appropriate independent practical state (or that they are a kind of hybrid cognitive/motive propositional attitude state).

At the very least, it does seem true that normative beliefs are *often* accompanied by the appropriate motivation. It may very well be that they are always so accompanied.

---

[162] Recall again that according to the orthodox interpretation (that of Rawls, for example) Hume himself would not have thought that there is such a thing as a normative belief. This interpretation has been challenged by more recent Hume scholars like Rachel Cohon. I will set such exegetical questions aside. For the purposes of my discussion here, it is sufficient to point out that remaining consistent with the three claims made by Humean psychologists does *not* require being consistent with Hume's own metaethical views. See Cohon (2008).

Nevertheless, as stated in the last chapter I am more sympathetic to the externalist intuition that weakness of will is possible—that the sincere amoralist is a character who could conceivably walk among us. Unfortunately, I will not be investigating these questions further. Rather, my goal in this chapter is to identify another sort of mental episode—one that even the Humean can count as hybrid. It is for this reason that I turn to the exploration of inner speech.

Part II: What is Inner Speech?

To begin our discussion of inner speech, I believe it will be helpful to first distinguish it from two other sorts of phenomena for which it might easily be mistaken. The first is the "mentalese" posited by various Language of Thought hypotheses. The second is what developmental psychologists refer to as "private speech."

Section 1: "Mentalese" and Cognitive Phenomenology

If Language of Thought theorists are right, then "speaking" in the language of thought (i.e. *thinking*) is a form of language use that is undertaken silently. To this extent there is a similarity between LoT and inner speech. However, there are two very significant respects in which LoT is nothing like inner speech. The first has to do with the fact that LoT constitutes a representational system that is itself essentially *not public*. It is entirely separate from natural language, having a syntax and semantics all its own. Inner speech, while often abbreviated and idiosyncratic, takes place in natural language.

The other point of contrast has to do with the fact that the Language of Thought Hypothesis is a version of the computational theory of mind. As such it is primarily a commitment to a particular view of the nature and structure of propositional attitudes: that the mental representations which are the objects of propositional attitudes belong to a symbolic system that admits of semantic and syntactic analysis. It is for this reason that LoT is sometimes also called *Mentalese*.[163]

In contrast, work on inner speech is not directly concerned with propositional attitudes as such. Inner speech refers to a subjective activity or experience that is episodic, not to psychological dispositions like propositional attitudes. Significantly, this means that even the most vocal skeptics of cognitive phenomenology in the case of propositional attitudes (those who doubt that beliefs, for example, have a distinctive "what it's like-ness") readily point to inner speech as a paradigm example of an inner process with a rich and accessible phenomenology.[164]

Throughout this essay I will be remaining as neutral as possible with respect to the debates about cognitive phenomenology and the language of thought hypothesis. What I *will* rely on is the fact that we all do have recognizable experiences to draw from as we think about inner speech. We all know *what it is like* to speak silently, and this experience is markedly different from that of *having a belief* or being in some other propositional attitude state (even if there *is* a distinctive phenomenology to being in such propositional attitude states).

---

[163] Aydede (2015).
[164] Carruthers & Veillet (2011), see especially pp. 35–38.

Section 2: From Private Speech to Inner Speech

The other phenomenon we should distinguish from inner speech is *private speech*. The two identifying features of private speech are (a) that it is audible and (b) that the speaker is not addressing an audience (aside from, perhaps, themselves). Usually private speech occurs when one is by oneself, but it can also happen within earshot of others (as when I overhear my roommate speaking in the shower). Though the phenomenon was first described by Swiss psychologist Jean Piaget as "ego-centric speech," "private speech" is the label most often used today.[165]

Private speech is especially common in young children between ages 3 and 7. Piaget therefore hypothesized that it was a symptom of a particular stage of cognitive development that children go through during which they are unable to understand other points of view. Piaget thought that these "monologues" of young children did not play a significant functional role in cognition, except as a kind of precursor to social language. We must be careful here, as Piaget himself shied away from making generalizations and constructing theories based on the data he collected. However, as Russian psychologist Lev Vygotsky later wrote:

> Piaget…did not escape the duality characteristic of psychology in the age of crisis. He tried to hide behind the wall of facts, but facts "betrayed" him, for they led to problems. Problems gave birth to theories, in spite of Piaget's determination to avoid them by closely following the experimental facts and disregarding for the time being that the very choice of experiments is determined by hypotheses. But facts are always examined in the light of

[165] Private speech also includes cases in which the primary audience is imaginary. It may very well be that the "privacy" of private speech comes in degrees, as is perhaps illustrated by Piaget's examples of what he calls "collective monologue." See Piaget (1959), Chapter 1 especially p. 31. I pass over the particular layer of complexity that arises from trying to account for the phenomenon of collective monologue as it does not arise for inner speech.

> some theory and therefore cannot be disentangled from philosophy. Who would find a key to the richness of the new facts must uncover the *philosophy of fact:* how it was found and how interpreted. Without such an analysis, fact will remain dead and mute.[166]

In his own investigation of the relationship between thought and language, Vygotsky took the careful examination of Piaget's shortcomings as his starting point. According to Piaget's reluctant theorizing, "ego-centric speech" is primarily a sign of immaturity that fades away as the child develops. Though it often *accompanies* the very behavior it is about, in such cases Piaget believed that the speech was still merely a *byproduct*—the same process that produces the behavior produces the speech.[167] Accordingly on Piaget's view, most of the time this speech served simply as a stimulus; children engage in it because it adds to their enjoyment of whatever activity they are engaged in. In this sense ego-centric speech might "accelerate" activity by making it more pleasurable, but it also runs the risk of "supplanting it" by distracting the child or giving her the feeling as of having completed her task (when really she has not).[168]

At its most fundamental point, Vygotsky's departure from Piaget involves the place at which he posited social speech entering the developmental picture and influencing the subsequent direction of development. In the introduction to his translation of Vygotsky's primary work on the subject, *Thought and Language*, Alex Kozulin makes this divergence particularly vivid:

> In Piaget's view…the uniqueness of speech-for-oneself, which is incomprehensible to others, is rooted in the child's original autism [undirected behavior] and egocentrism, and ultimately in the pleasure

---

[166] Vygotsky (1986), p. 15.
[167] Vygotsky (1986), pp. 28–29.
[168] Piaget (1959), p. 33.

> principle. In the course of the child's development this individual speech *dies out, giving place to* socialized speech, which is easily understood by any interlocutor, and which is ultimately connected with the reality principle.
>
> Without denying the phenomenon of autism as such, Vygotsky suggested that egocentric speech is rather a transitory form situated *between* social, communicative speech and inner speech. For Vygotsky the major problem was not that of socialization, but rather of individualization of the originally communicative speech-for-others….Vygotsky believed that the outward, interpsychological relations *become* the inner, intrapsychological mental functions. In the context of this idea, the transition from egocentric to inner speech manifests the internalization of an originally communicative function, which becomes individualized inner mental function.[169]

This last paragraph is crucial. For Piaget the puzzle was to explain how a child could transition from an ego-centric form of speech, into full-fledged social interaction, awareness, and understanding. By redescribing private speech as the developmental step *between* social speech and silent inner speech, Vygotsky transformed the question to be answered. Rather than ask how the individual becomes social, Vygotsky claimed that we should be asking how the social could become individual. Because social speech is developmentally prior, it must be the case that somehow "the interpsychological relations *become* the inner, intrapsychological mental functions." This is why, according to Vygotsky, private speech plays a critical part in cognitive development and social learning.[170]

Vygotsky hypothesized that private speech functioned as a form of self-directed regulation that gradually became internalized. He relates a particular episode that occurred during his experiments that illustrates well the sort of evidence that led him to object to Piaget:

---

[169] Vygotsky (1986), p. *xxxvi*. Emphasis added.
[170] See Vygotsky (1986), Chapter 2 especially pp. 86–87.

> A child of five-and-a-half was drawing a streetcar when the point of his pencil broke. He tried, nevertheless, to finish the circle of wheel, pressing down on the pencil very hard, but nothing showed on the paper except a deep colorless line. The child muttered to himself, "It's broken," put aside the pencil, took watercolors instead, and began drawing a *broken* streetcar after an accident, continuing to talk to himself from time to time about the change in his picture. The child's accidentally provoked egocentric utterance so manifestly affected his activity that it is impossible to mistake it for a mere byproduct, an accompaniment not interfering with the melody. Our experiments showed highly complex changes in the interrelation of activity and egocentric talk. We observed how egocentric speech at first marked the end result or a turning point in an activity, then was gradually shifted toward the middle and finally to the beginning of the activity, taking on a directing, planning function and raising the child's acts to the level of purposeful behavior.

In contrast with Piaget's hypothesis that such speech is a mere immature epiphenomenon, Vygotsky's understanding of private speech was that it is an essential aspect of cognition that never really goes away.[171] Rather, it just gradually becomes silent. "Inner speech" is the term used today to pick out this silent, mature, and fully internalized form of private speech.[172]

At the center of Vygotsky's theory was a hypothesis that also forms a central hypothesis of this chapter. Here is Kozulin again:

> Peculiarities of grammer and syntax characteristic of inner speech indicate this submergence of communication-for-others into individualized reasoning-for-oneself: in inner speech, culturally prescribed forms of language and reasoning find their individualized realization. Culturally sanctioned symbolic systems are remodeled into individual verbal thought. The principal steps in this remodeling include the transition from overt dialogue to internal dialogue. [173]

---

[171] Tappan (1997), pp. 80–81.
[172] Vicente & Manrique (2011).
[173] Vygotsky (1986), p. *xxxvi*.

The original use and function of natural language is social; the interpersonal sphere is the one in which language finds its genesis and significance. The Vygotskian hypothesis is that inner speech is the end result of a process of internalizing public, interpersonal speech. Understanding these social roots is the key to understanding inner speech's structure and the functional role that it is able to play in cognitive processes. In this chapter I hope to enrich this story by pointing to the pragmatic dimension. Not only are the syntax and semantics of inner speech derived from its social origin in natural language, the constitutive norms, expressive power, and psychological function of inner *speech acts* are also derived from the corresponding pragmatic features of *social* speech acts. So on the one hand this chapter argues *from within* the Vygotskian framework, but I also take it to provide an argument *for* that same framework. That is to say, I begin from Vygotskian assumptions about the developmental picture of inner speech, and argue that essential characteristics of speech acts discussed in the last chapter support that very picture.

Inner speech is that activity we are engaging in when we speak silently in a natural language—typically (though not always) our first language. Usually this silent speech is idiosyncratic and abbreviated compared to the way we would communicate publicly with another person—still, it is decidedly an internalized version of natural language. In the same way that your conversations with a partner or close friend, though built out of natural language, might be difficult for a stranger to understand, so too (and even more so!) it is often the case in our conversations with ourselves. Nevertheless, even the most condensed intra-personal speech may be expanded back into the form that inter-personal speech

takes—in fact, if Vygotsky's followers are correct, the former is ultimately parasitic (both structurally and developmentally) on the latter.[174]

Section 3: Inner Acts of Speech vs. Inner Speech Acts

In his book *Self Expression* Mitch Green outlines a distinction that has been and will continue to be crucial to our discussion. This is the distinction between "acts of speech" and "speech acts":

> Factual speaker meaning usually occurs in the context of a speech act. Let me make clear that speech acts are to be distinguished from acts of speech. When I test a microphone, utter lines on stage, or practice a speech in the shower, I am performing acts of speech but no speech acts. By contrast, a speech act is any act that can be performed by, under the right circumstances, speaker-meaning that one is doing so. I can raise a question ('I ask you what time it is'), make a statement ('I state that it is 5 p.m.'), issue a command ('I command you to make the appointment on time') by saying that I am doing so in such a way as to speaker-mean it. This is why questioning, stating, and commanding are speech acts. I can also perform one of these acts by speaker-meaning a content in a certain way but without saying that I am doing so. I can assert that snow is white without saying that I am asserting that snow is white. I can do that simply by uttering the words, 'Snow is white', meaning this as an assertion.[175]

As in public speech, so also in inner speech. When we speak silently to ourselves, sometimes we perform mere inner *acts of speech* while at other times we perform inner *speech acts*.

---

[174] For Charles Fernyhough's Vygotskian characterization of the structure of this internalization process, as well as more on the difference between expanded and condensed inner speech, see Fernyhough (2004).

[175] Green (2007), pp. 69–70. Vygotsky seems to have something like this distinction in mind when he talks about "nonintellectual speech"—see Vygotsky (1986) p. 88–89.

For example, it is often the case that when individuals read (especially if they are reading carefully) they will repeat the words on the page silently as they read them.[176] You may be doing this yourself as you read this chapter. This silent articulation is merely an inner act of speech. As you silently utter the words you are reading, you don't *mean* anything by so uttering. However, suppose that in the course of reading you stop to silently pose yourself a question—*"Am I really doing what the author is describing?"*—then you have performed an inner speech act. Or again, perhaps you need to remember a phone number you have just been told while you search around for a pencil and paper to write it down. Your silent repeating of the digits over to yourself as you scrounge for writing implements is a perfect example of the phonological loop component of working memory. It is an inner act of speech.[177] But when you pause to say a silent promise to yourself—"*I won't forget to call them back*"—this is an inner speech act.

In the psychology literature, inner speech that is composed of inner *acts of speech* is part of the broader category of "subvocalizations," while the subset of inner speech composed of inner *speech acts* is part of a category that is referred to as being "dialogic."[178] Over the past 15 years the importance of distinguishing the dialogic sphere of inner speech from non-dialogic subvocal articulation has become more and more appreciated. For example, Simon Jones and Charles Fernyhough posit in their survey of the literature on the relationship between the neural correlates of inner speech and auditory verbal

---

[176] Morin (2012) pp. 440–441. For more on the role of inner speech in reading see Ehrich (2006).

[177] This is the area of inner speech focused on by Alan Baddeley and Graham Hitch in their phonological loop model of working memory. See Baddeley and Hitch (1974).

[178] For example see Winsler (2009) and Morin (2012).

hallucinations that carelessness with respect to the difference in cognitive significance between dialogic inner speech and mere subvocalization has been a source of confusion in the investigation of the target phenomenon.[179]

In that essay, Jones & Fernyhough outline some very preliminary findings that, if substantiated would suggest that different kinds of inner speech are associated with higher and lower cognitive loads on the Verbal Self-Monitoring system.[180] Specifically, *dialogic* inner speech (as in making statements to oneself or imagining statements being made) could involve higher levels of VSM, while *subvocal articulation* could involve lower levels (as in silent reading or the repetition characteristic of working memory's phonological loop component). Here's how one would describe the possibility in terms of the pragmatics concepts from the last chapter: when we perform speech acts silently to ourselves (or imagine them being performed) we engage in a higher level of self-monitoring, and when we perform silent acts of speech we engage in less self-monitoring.

I do not want to make too much of the empirical data. This is an area on which significantly more work needs to be done before conclusions can be drawn. In particular, the category of dialogic inner speech needs to be divided into the separate activities of silently *performing* inner speech acts and silently *imagining* inner speech acts being performed. For our purposes it is these distinctions that are important, as they clearly

---

[179] Jones and Fernyhough (2007).

[180] As background, it is helpful to know that one of the leading hypotheses concerning the neural correlates of inner speech is that it involves activation both of the left inferior frontal gyrus/Broca's area and the right temporal cortex. What enables us to recognize inner speech as our own is the verbal self-monitoring (VSM) system. The VSM mechanism functions at least in part as a system of corollary discharges: when we engage in inner speech, the speech producing areas of the frontal lobes send dampening signals to the auditory cortex.

identify the sort of evidence for which we should be on the lookout. The fact that further work *could* support a general difference in the cognitive significance of dialogic inner speech as compared with mere subvocal articulation is important. And such a discovery would not be surprising. In a natural parallel, the performing of a public speech act has quite a different level of social significance than does a mere public act of speech. As we saw in the last chapter, speech acts—because of the norms that constitute them—enable us to do quite a wide variety of things.

For example, by making a promise to you I can *undertake a commitment*. By making an assertion I can *express a belief*. By christening a ship, I can make it the case *that the ship has a certain name*. By apologizing I can *express my remorse*. By marrying a couple, I can bring into existence particular kind of relationship between two people. Many speech acts, in order to be successful require the person performing them to inhabit a particular sort of position within a particular sort of context. Not just anyone can excommunicate you from the Roman Catholic Church, but those who can are such that *their saying makes it so.* Other speech acts like asserting, apologizing, promising, conjecturing, etc. allow virtually anyone to perform particular actions merely by intentionally saying the right sort of thing with the right sort of pragmatic force (e.g. assertive, apologetic, promissory, conjectural, etc.).[181]

What happens when we perform speech acts in the privacy of our own thoughts—when we perform *inner speech acts*? Clearly there are many speech act types

[181] One is tempted to add "and to the right sort of audience." That is to say, one common sort of criteria for speaker meaning is *reflexive communicative intention*. Rather than ascend Grice's ladder myself, drawing out the difficulties along the way, I will simply direct the reader to Green (2007) chapter 3. As we will see later, it is important that this condition is *not* required.

(christening a ship, for instance) that could not possibly occur with full felicity in the privacy of one's own thoughts. Is this true of all speech act types? If some speech act types may felicitously be performed silently to oneself, how do they inherit the norms of their social counterparts such that they may be judged to be speech acts of *the same type*? After a survey of recent empirical investigations of inner speech in Part III, I will return to these questions about the nature of inner speech acts in Part IV.

Part III: Current Work on Inner Speech

In what follows I give a brief overview of some current research on inner speech, and an argument for the possibility of inner speech acts. I highlight the influence of Vygotsky's framework, as well as its explanatory power in the realm of dialogic inner speech in particular. One goal of this discussion is to demonstrate the connections between social speech and inner speech, including the way these connections help to explain the functional roles that inner speech is able to play. However, my primary goal is to show that distinguishing between dialogic inner speech (especially inner speech acts) and non-dialogic subvocal articulation (inner acts of speech) is essential to our understanding of and continued research into the phenomenon of inner speech. In Part IV I will turn to the issue of whether the possibility of inner speech acts is consistent with the status quo in general speech act theory. I argue that the theory of speech acts employed in the last chapter is compatible with the existence of inner speech acts. This approach to speech act theory, sometimes referred to as force conventionalism, is also the *only* perspective that does not

seem to rule out the possibility of inner speech acts from the start.[182] Moreover, Force Conventionalism and the Vygotskian theory of inner speech are mutually supporting; each helps to illuminate the other. In Part V I will return to the normative dimension to trace the implications of a newly minted Vygotskian hybrid inner speech act theory for our normative mental life.

Section 1: Current Research on Inner Speech

In his essay "Dialogic Thinking," Charles Fernyhough's goal is to investigate the properties of inner speech that could make it particularly well suited to play a substantial role in cognitive processes. He begins by defining what it takes for a mental process to count as "dialogic" more generally:

1. The process involves simultaneous accommodation of multiple perspectives on reality
2. These perspectives are represented in an interpretable sign system (like natural language)
3. The perspectives are flexibly coordinated
4. The perspectives preserve the triadic intentional relations of external dialogue
5. The perspectives interact with each other in a way that is open-ended and self-regulating[183]

What does Fernyhough mean by "accommodation"? Because his goal is primarily to investigate the phenomenology of inner speech, we should not read Fernyhough as making

[182] The label "force conventionalism" to describe this view comes from Green (2015).

[183] Fernyhough (2009), p. 43.

a deep metaphysical claim about the number of interlocutors that must literally be present *inside* a person in order for them to engage in dialogical inner speech. Rather, Fernyhough's "multiple perspectives" refer to what an individual must *represent* as she speaks dialogically. Flexible coordination consists in the ongoing modification of these perspectival representations by their dialogically represented counterparts (e.g. question/answer, assertion/disagreement). In inner speech this involves the thinker's treating of each perspective *as if* it had come from another person, and responding accordingly.

The triadic intentional relation represents the way in which two or more interlocutors are related both to one another (qua interlocutors), and also some other element of reality (the object of their discourse). Just as the external dialogue between two individuals requires no direction from a third source, so too there is no need to postulate a superordinate supervisory system responsible for directing the flow of the internal process when it possesses this dialogic character (open-ended and self-regulating). Fernyhough argues that when inner speech displays these qualities, it creates a unique opportunity for thinkers to represent the world (and their attitudes towards the world) *for themselves* in a way that reduces processing cost and streamlines working memory.

This description of what we are doing in inner speech may sound strange at first. Most of the time when we speak silently, even when such silent speech is dialogic, it does not consist in question, answer, assertion, negation, etc., in the same sense in which public conversations sometimes do. In the non-pathological case, I would not silently ask a question as a *query* to gain information. I do not argue with myself by asserting, with full confidence, contradictory claims. Is it misleading to describe quotidian dialogic inner

speech as "conversational"? I don't think so, for in fact there are appropriate analogues in social speech for the sorts of conversations we represent in inner speech. Oftentimes our conversations with others are not guided by goals like transferring information or contradicting one another. In many cases our goal is to solve a problem, express how we feel, decide how to act, determine what to believe, or explore an idea. In such conversations—much like in inner speech—questions are posed to direct inquiry, assertions are made to make explicit relevant information or commitments, claims are put forward as hypotheticals for consideration, perspectives are tried out and accepted or rejected. The misconception that our conversations with one another are always about communication (understood narrowly as *the mere exchange of information*) can easily lead to skepticism about truly dialogic inner speech. This is an error that we will return to later as well.

For the purposes of my project here I will describe as "dialogic" all inner speech that possesses a conversational structure. This includes all inner speech acts (asserting, questioning, wondering, etc.) whether performed earnestly by the thinker in question or imagined by them to be performed by another. Thus silently rehearsing or inventing a conversation in which I am not one of the interlocutors would still count as a dialogic inner speech, even though from the perspective of pragmatics it would not involve genuine speech act performance. I believe that this comes closest to capturing the way the term 'dialogic' is actually used by psychologists like Fernyhough. Thus while all dialogic inner speech is importantly distinguished from subvocal articulations that do not possess a conversational format, it is the proper *subset* of dialogic inner speech that consists in genuine speech acts that will prove most important for my investigation.

Now let us turn to the advantages that come from adopting Vygotsky's approach to inner speech. Fernando Martínez-Manrique and Augustin Vicente defend the Vygotskian approach against the inner speech theories of José Luis Bermúdez, Jesse Prinz, and Peter Carruthers, by pointing out its unique ability to explain the diversity of forms that our inner speech can take.[184] Even if we restrict ourselves to discussing particularly *dialogic* inner speech, such discourse can still come in both condensed and expanded varieties:

> [Condensed] inner speech not only lacks pitch and volume but also appears typically in subsentential linguistic items. So [this form of] inner speech is not experienced in the way of the internal monologues that classical novels depict but in the fragmentary way exemplified by the opening of Samuel Beckett's "The Unnamable": "Where now? Who now? When now? Unquestioning". Our "outer talk" also often has subsentential utterances but our inner talk often seems to be full of expressions like 'ah', 'yes', 'not this way', 'where the hell?' and 'the meeting!' (see Peacocke, 2007). The meaning of those linguistic items is typically clear to us but they might be to a large extent unintelligible to others if we uttered them (see on this respect the transcription of writers' notebooks in John-Steiner, 1997: 111ff).[185]

On the other hand:

> [Expanded inner speech is] more "sophisticated" inner talk, seemingly carried out in full sentences. This is especially noteworthy in cases such as when we prepare a lecture, think hard about an argument, or imagine possible conversations. Many of those cases seem to be related to linguistic actions, i.e., what our inner speech is doing can be characterized as a sort of rehearsal of the utterances that the subject will eventually make public. However, sophisticated inner speech may also take place in other kinds of situations. For instance, in a research comparing the phenomenological qualities of inner speech in voice-hearing schizophrenia patients and healthy controls, Langdon et al (2009) found that both groups were most likely to report thinking in full sentences.[186]

---

[184] See Bermúdez (2003), Prinz (2000 and 2007), and Carruthers (2006).
[185] Martínez-Manrique & Vicente (2010), p. 2.
[186] Martínez-Manrique & Vicente (2010), pp. 2–3.

The majority of inner speech is condensed. On the Vygotskian model, this makes sense. If we understand inner speech as an internalized form of social speech, then it will operate by using analogues of the same principles at play in social speech. Martínez-Manrique & Vicente point out that one of the mechanisms that enables public, social speech to be condensed is *common ground*. When two interlocutors are engaged in conversation, there is a certain amount of information that is understood by both parties to be shared and available between them. This allows for the ease with which we can make use of linguistic tools like indexicals, presupposition, implicature, ellipsis, etc.

For example, if the two of us are having a conversation when all of a sudden the wind pushes the door open with a loud bang, you may say "that was startling!" without having to explicitly clarify that you meant to use 'that' to refer to the event of the door opening. The immediacy and force of its occurrence enters the event into our common ground as a salient object of reference, allowing you to easily direct my attention towards it (you don't even have to point). Whenever two people share a particularly large and intricate array of information in common ground, extremely clear, abbreviated, and efficient conversations are possible. These conversations are also extremely difficult for outsiders to enter into. One need only eavesdrop on a couple with a long and healthy marriage, soldiers from the same military unit, athletes on the same team, or experts from the same academic field to quickly get the sense that one is missing out on the most important aspects of their conversations.[187]

[187] Vygotsky himself makes a similar connection (though he is unable to refer to the concept of "common ground") in his discussion of the idiomatic quality of inner speech. See Tappan (1997) pp. 82–83. It is important that I describe what you are doing

According to the Vygotskian perspective, inner speech takes on a condensed form for a similar reason. Despite inner speech's subsentential structure it is still a fully dialogic process, it's just that the role that would have been played by common ground becomes unnecessary. There is no epistemic distance between speaker and audience: the relationship between the two is one of identity.[188] The "common ground" we share with ourselves includes *everything* that is going on in our heads—perceptions, sensations, mental imagery, occurrent propositional attitudes, the accessible contents of long and short term memory, etc. This allows our conversations "with ourselves" to exhibit a significantly abbreviated structure—a level of efficiency and transparency that is unattainable in social speech.[189] In contrast, because expanded inner speech is often serving as a form of preparation or rehearsal for *social* speech, the same level of subsentential abbreviation would be counterproductive.

Like Charles Fernyhough, Fernando Martínez Manrique and Agustin Vicente see the dialogic quality of inner speech and its Vygotskian origin in social speech as being the keys to understanding inner speech's substantial cognitive role. They summarize their conclusion clearly as follows:

> In a nutshell, language is claimed to be the means by which we can "objectify and contemplate our own thoughts". Converting thoughts in

---

as "eavesdropping." Were you to be a fully-fledged member of the conversation, the way of speaking would most likely be adapted to account for the smaller area of common ground (at least, that would be the polite thing to do).

[188] Which is why I allow for the possibility that distinguishing a "speaker" from an "audience" in such cases is a merely metaphorical indulgence.

[189] This could at least partly explain why subjects report speaking "impossibly fast" in inner speech. When describing to a researcher what it is that they said in an episode of condescend inner speech, the subject may expand the utterance in question without noticing they are doing so. Successfully expressing the same proposition in social speech takes longer, because it takes more phonemes. See Fernyhough (2016) Chapter 5.

> objects by formulating them linguistically enables us to hold the focus on our thinking, which in turn enables us to have a better control of our behaviour, of our planning and of our cognitive processing in general. It is not that language allows us to have second-order thoughts; rather, the idea is that language allows us to make thoughts conscious. We can be conscious of our thoughts through language, just as we are able to express our thoughts by language, and neither of those abilities requires that linguistic representations carry the thoughts themselves, i.e., it is not necessary to suppose that the representational format of thoughts is natural language itself. To put the idea in other terms: it is not that we recruit "language", whatever that is, to play a "second-order dynamics", as Clark (1998) calls it. Rather, we recruit linguistic communication, that is, we convert a pattern of outward actions, into a form of cognition. When we hear our inner voice we are listening to ourselves talking to ourselves. We recruit the way we have to tell our thoughts to others in order to tell them to ourselves (with all the cognitive changes this brings in).[190]

This is a strong conclusion, and my purpose here is not to defend it. For one thing I find it somewhat misleading to describe *hearing the inner voice* as "listening to ourselves talking to ourselves." This way of speaking invites positing a stronger division within the individual than I am comfortable with, at least absent abnormal psychological conditions like dissociative identity disorder. However, inner speech *is* an important tool for facilitating self-awareness, even if it involves speaking to no audience in particular. Without a distinction between dialogic inner speech acts and mere silent acts of speech, it would be impossible to explain this functional role it plays in consciousness.

Alain Morin argues in this same vein for the Vygotskian perspective's ability to uncover the link between inner speech and self-awareness.[191] In his 2005 essay he proposes three main sources of self-awareness: the social environment (face-to-face interactions,

---

[190] Martínez-Manrique & Vicente (2010), p. 4.

[191] For an overview of his work, see Morin (2012). For a defense of the Vygotskian approach emphasizing the self-regulatory function of inner speech and the importance of attention control, see Clowes (2007).

reflected appraisals, etc.), the physical world (mirrors, written material/media, etc.), and the self (proprioception, imagery, etc.). Morin posits inner speech as a process that allows the self to take tools from the social milieu and the physical environment and use them to enhance self-aware cognitive reflection.

With respect to the social milieu, inner speech enables us to reproduce a variety of social mechanisms in a way that facilitates self-awareness. For example, inner speech can allow you to replicate comments emitted by others, or internalize others' perspectives. By initiating a silent dialogue, inner speech lets you describe a situation or explain yourself in response to the reactions of an imagined audience. Extending the scrutiny of others by engaging in inner speech, you can recreate the sense of being actively observed. This is a process that fosters self-focus.[192] Sometimes this means imagining a conversation in a way that does not commit oneself to a particular perspective—trying out points of view in a risk-free way.[193]

Morin goes further: not only does inner speech allow for perspective taking, it also enables self-representation. When a person talks *about* an inner state ("I'm really cold!", "I feel disappointed…", "How exciting!"), such an activity can facilitate bringing information about the self to the level of conscious awareness. Such a person is not only *in* the state, they *know* that they are in the state. By verbally labeling self-aspects, inner speech allows one to become aware of those self-aspects (emotional responses, physiological sensations, values, attitudes, goals, etc.).[194] It is this aspect of the process of

---

[192] Morin (2005), pp. 122–123.

[193] For the purposes of precision and clarity, remember that this activity of imagining speech acts being performed (though dialogic in form) would fall short of making genuine inner speech acts.

[194] Morin (2005), pp. 125–126.

self-awareness that Adam Smith is gesturing towards in the excerpt quoted at the beginning of this chapter.[195]

For empirical evidence, Morin cites both neuroimaging studies and analyses of self-reports. Of course, there is substantial debate about precisely which areas of the brain are implicated in self-awareness. However there does seem to be something of a consensus on the hypothesis that, at the very least, portions of the left prefrontal lobe are associated with both self-awareness and inner speech.[196] Some of the most compelling support for this thesis comes from cases in which a patient has suffered damage to that part of the brain and exhibits corresponding self-awareness and inner speech deficits. Morin offers the example of the clinical psychologist C. Scott Moss who recuperated from aphasia after suffering a stroke. Here's how Moss relates his experience:

> The second week [at the hospital] I ran into a colleague who happened to mention that it must be very frustrating for me to be aphasic since prior to that I had been so verbally facile. [I] later found myself why it was not. I think part of the explanation was relatively simple. If I had lost the ability to converse with others, I had also lost the ability to engage in self-talk. In other words, I did not have the ability to think about the future—to worry, to anticipate or perceive it—at least not with words. Thus for the first four or five weeks after hospitalization I simply existed. So the fact that I could not use words even internally was, in fact, a safeguard.[197]

---

[195] One is reminded here of feminist and Marxist analyses of the ways that power relations can deny marginalized groups even the ability to describe their experiences. The act of naming or labelling an experience can be a powerful moment of self-discovery, whether that act is performed silently or aloud. See Fricker (2007), Chapter 7.

[196] Morin (2005), pp. 126–127. For a careful untangling of the complexity added to the debate by conflicting neurocognitive theories of the *levels* of consciousness and self-awareness, see Morin (2006).

[197] Moss (1972), p. 10.

It is not just the ability to silently play out imaginary conversations that Moss describes as lost. More importantly, it is the ability to perform inner speech acts—speech acts that could serve to express inner states like worries, anticipatings, and perceptions.

In addition to self-awareness, inner speech also plays a substantial role in action. For example, Annalisa Sannino argues that Vygotskian insights are crucial to closing the gap between discourse and activity.[198] As discussed above, it is crucial to Vygotsky's theory that conversations (both social and private) are not only products of activity, but also *generators*. By introducing Vygotsky's understanding of inner speech as dependent on and derived from social speech, Sannino enriches Margaret Archer's "internal conversation" contribution to the structure-agency debate. This longstanding discussion in social theory concerns sorting out the influences of structural forces present in an agent's environment and carving out a role for the agent's own independent capacity to act and make free choices. Archer theorizes that in the very process by which structure influences the agent, the agent herself acts as a medium that shapes and determines that influence. Internal conversation, a mental activity by which the agent renders the external world intelligible, represents a crucial aspect of this shaping process.

The weakness of Archer's view comes from her tendency to see the agent as isolated. By drawing on Vygotskian observations, Sannino recasts inner speech as itself an essentially social phenomenon.

> The determination of the subject to wield individual power stems from the property of the agent to converse internally and deliberate upon the social circumstances. Archer's definition of internal conversation is an abstract and individualistic property of a subject who is more or less capable to effectively act in the environment. Agency is an active orientation toward

[198] Sannino (2008).

> the external world. As long as internal conversation remains a category confined within the individual subject it cannot fulfill its agentive potential. This potential lies in the fact that internal conversation also inherently implies material externalization in dialogue with others and movement toward material actions.[199]

If we were to confine our understanding of inner speech to instances of subvocal articulation and acts of speech, we could not explain the role that it plays in action. It is precisely because inner speech can take on the characteristic weight of social speech—*committing* oneself to perspectives and plans—that it vitalizes our agency. Once again, dialogic inner speech (particularly in the form of inner *speech act*) is essential to capturing the full range of our experience.

Charles Fernyhough (2008) explores the details of the psychological mechanisms involved in inner speech's influence on agency by examining the development of social understanding. In Fernyhough (2010) he expands this into a Vygotskian theory of executive functioning. The primary challenge here is to explain the emergence of social understanding without falling into either an unsatisfying form of dualism or a crude reductive behaviorism, and to do so in a way that respects the influence of the social environment on executive functioning. Fernyhough's solution is to develop a Vygotskian "interfunctional" approach that identifies childhood linguistically mediated social experiences as the third factor which determines development of both social understanding and executive functioning. He predicts that "if mediated EF [executive functioning] is derived from social interaction, as Vygotsky's theory holds, then it should be possible to observe influences of social interactional experience on private speech and EF

[199] Sannino (2008), p. 288.

development." So far the empirical research has been consistent with Fernyhough's hypothesis. For example, studies have shown correlations between adult-child interaction and subsequent private speech development. In one case this involved relating to culturally specific experiences, while in another it linked deficits in reciprocal interactions to delays in private speech development.[200]

The key to the solving puzzle is inner speech, specifically as exhibited in its dialogic mode:

> Viewing both EF and SU [social understanding] as mediated higher mental functions thus presents an alternative way of conceptualizing the developmental relations between these variables. Rather than attempting to account for the relation in terms of a strong executive component to SU tasks, or conversely an SU load on standard EF tasks, a combined interfunctional approach to EF and SU would see the development of both capacities as being driven by *the internalization of dialogic, mediated interpersonal activity*.[201]

One form of inner speech Fernyhough is referring to here involves internalizing alternative perspectives, an activity whose link to social understanding would be unsurprising. However, Fernyhough takes it to be essential that "progress in SU development should go hand-in-hand with a semiotic shift in other domains."[202] Rather than explaining the connection between them by analyzing either SU or EF as more fundamental, Fernyhough's theory the links the two processes by uncovering their shared dependence on the capacity for internalized dialogic activity. This is what qualifies both as "mediated higher mental functions." For example, just as Social Understanding increases with our

---

[200] Fernyhough (2010), p. 65.
[201] Fernyhough (2010), pp. 68–69. Emphasis added.
[202] Fernyhough (2010), p. 68.

ability to silently imagine conversations, the development of Executive Functioning is facilitated by self-regulatory private speech. SU and EF both depend on our capacities for private and inner speech, and EF specifically makes use of forms of solitary talk that rise to the level of speech act.

Mark Tappan takes a similar approach to explain moral functioning in particular. For Tappan, inner speech is central to understanding moral development, deliberation, and action. Here are the four core theses of his Vygotskian informed sociocultural perspective:

> 1) moral functioning (like all "higher psychological functioning") is necessarily mediated by words, language, and forms of discourse;
>
> 2) such mediation occurs primarily in private or inner speech, typically in the form of inner moral dialogue;
>
> 3) because language is the social medium par excellence, processes of social communication and social relations necessarily give rise to moral functioning;
>
> 4) because words, language, and forms of discourse are inherently sociocultural phenomena, moral development is always shaped by the particular social, cultural, and historical context in which it occurs.[203]

While I am hesitant to unreservedly endorse all these claims myself, Tappan does make a strong case that inner speech plays an essential role in a great deal of our moral functioning. Take for example his emphasis on the moral point of view's origin in shared practices:

> [I]t is these shared activities that enable persons to understand the predicatized forms of speech by means of which they communicate about moral issues with themselves and with each other- if they did not share these activities they would not share a common moral language. While Vygotsky's (1934/1986) example of communication carried on by means of a common vernacular language that emerges from shared activity is somewhat more mundane—a conversation, reported by Dostoevsky,

[203] Tappan (1997), p. 83.

> between six drunken workmen that consisted entirely of the same single epithet, repeated in different tones and inflections—it nevertheless illustrates quite well the process by which common, everyday words (in both inner and outer speech) mediate thought, feeling, and action.[204]

This idea, that vernacular moral language mediates and shapes moral functioning has been supported by substantial empirical research.[205] For example, studies of American moral discourse (both social and solitary) have identified various distinct moral "voices" with which an agent may speak, each of which is characterized by a certain vocabulary. One prominent such voice promotes the interests of *justice* through deontic language like 'right', 'harm', and 'duty'; another advocates for considerations of *care* using terms like 'love', 'hurt', or 'vulnerable'.[206] While most individuals make use of multiple voices, sometimes even vacillating between them in the course of a single deliberation, one particular voice is often dominant. Other secondary "voices" have also been studied, such as those that incorporate perspectives and language appropriated from Christianity or Republicanism.[207] In contrast, compare studies of traditional Chinese moral discourse that is shaped by vocabulary concerned with ethical concepts like filial piety.[208] Or consider a revealing study of Tibetan Buddhist monks that demonstrates the inadequacy of a moral reasoning analysis scheme when applied to a moral community that is different from the one in which the diagnostic tools were originally developed.[209]

---

[204] Tappan (1997), pp. 85–86.
[205] For an overview, see Tappan (2006a) and (2006b)
[206] Gilligan (1982), Gilligan (1983), Gilligan & Attanucci (1988).
[207] Bellah et al. (1985).
[208] Dien (1982).
[209] Huebner & Garrod (1993).

Ultimately, Tappan presents a picture of the moral self as dialogic, and of healthy moral communities as those that foster open conversations. Drawing on the work of Carol Gilligan and Mikhail Bakhtin, Tappan argues that "moral thoughts, feelings, and actions are semiotically mediated, and thus socioculturally situated," and hence the most natural forms of moral deliberation are narrative and discursive.[210] In attempting to understand moral psychology, Tappan thinks our focus should be turned more towards interpreting self-reports and conversations. In attempting to determine the right thing to do, our focus should be on engaging in moral conversation, both with ourselves and with others.[211] I will return to this idea of the dialogic moral self later in the chapter. First we must investigate the implications of speech act theory for our understanding of dialogic inner speech in general.

Section 2: A Brief Digression on the Two-in-One

In her unfinished work *The Life of the Mind*, Hannah Arendt offers a theory of thought inspired by Socrates in which the idea of conversing with ourselves is taken seriously. In the chapter titled "The two-in-one" the picture Arendt paints is decidedly that of a duality of self. The activity of thinking is not just one in which we *make as if* to take on different roles for the sake of deliberation, but in which two different perspectives are truly present. These two perspectives may sometimes be at odds, but for thinking to take place they must at least be on friendly terms. As Elizabeth Young-Bruehl puts it:

> Thinking was, for Arendt, "dialectical." This "frozen concept" she resolved into its original meaning: "the soundless dialogue [*dialegisthai* as 'talking

[210] Tappan (2006a), p. 352.
[211] Tappan (1990) and (2006b).

> through words'] between me and myself" (I, 185). The actualization of our internal plurality has the effect of liberating us not only from conventional "truths" but from conventional rules of conduct. The last is, of course, crucial when conventions of conduct, to take one example, make murder the "norm." What a thinking person will not do is live with a murderer, live with an internal "myself" who is a murderer. From the point of view of the "thinking ego," it is better to suffer wrong than do wrong and live with the wrongdoer.[212]

It is not always entirely clear whether, for Arendt, the two-in-one is meant to be taken literally. At one point she suggests that it is meant only "metaphorically and tentatively."[213] However, others *have* taken the idea more literally and it is that suggestion that I would like to address.

Stephen Darwall and Christine Korsgaard claim that part of what it is to be a rational being is to stand in a "second-personal relation" to oneself. Specifically, we are said to "interact with," "make demands on," and "obligate" ourselves in *exactly the same way* that we are able to stand in these relations with others.[214] Matthias Hasse has objected that such a view "rests on a logical confusion."[215] Essentially his objection is that any analysis of the second-person pronoun that allows reflexive and non-reflexive uses to bear the same logical structure will be unable to support the possibility of mutual recognition. If Haase is right, then this would be a problem for philosophers like Korsgaard and Darwall *precisely because* they want this second-personal relation to serve as a foundation for mutual recognition, with our relationship to ourselves both modeling and grounding the normative demand to respect others. Haase argues that for mutual recognition to be

---

[212] Young-Bruehl 1982, p. 279.

[213] Arendt (1961), p. 13.

[214] See Darwall (2006) especially Chapter 1, Korsgaard (2007), and Korsgaard (2009) especially Chapter 9.

[215] Haase (2014), p. 358.

possible, the act of address must be an essentially linguistic *sui generis* act of mind—one that it is by definition impossible to direct towards oneself.

Does my understanding of dialogic inner speech in general, and of inner speech acts in particular, assume that a person must be able to literally speak *to* themselves—to be their own audience? In what follows I strive to give an account of inner speech (including inner speech acts) that is neutral with respect to this issue. Some of the authors I surveyed in the previous section at times certainly write as though they share Korsgaard and Darwall's perspective. Mostly I will let this way of speaking pass as merely metaphorical. It is very difficult to describe what we are doing when we think without making reference to the representation of two or more different perspectives, especially given that inner conversations often involve multiple distinct voices.[216] To make the situation somewhat clearer, it is helpful to distinguish between phenomenology and metaphysics. When researchers on inner speech describe our many voices speaking to one another they are describing *what it is like* to engage in self-talk, not counting how many subjects exist.

At the same time, my own view about inner speech acts does require that silent speakers often *actually do* perform speech acts—it does not merely *seem to them* as though they do. This means that we need to be able to make sense of the idea of performing a speech act, and of speaker meaning in general, in the absence of addressing an audience. Haase is skeptical of the prospects for such a view. His worry becomes especially pressing in the case of speech acts like promising:

---

[216] Consider for example the inner speech and voice hearing experienced by authors. For many novelists and other writers, creating a character for a story often involves creating a distinctive voice in their own inner life. See Fernyhough (2016), Chapter 12.

> For the obligation to be created, the promise must be accepted by the addressee. It is a condition that the recipient, Y, wants what X promises to do. This entails that Y can release X from the obligation if she does not want it anymore. Otherwise X's act of promising would bind not just X's, but also Y's will. If you could not release me from the obligation I have to you in virtue of promising to cut your hair tomorrow, my promising would obligate you to getting your hair cut by me. Of course, we might promise each other to have a haircutting session tomorrow. But if my promise could bind your will, your acceptance would not be required in the first place. Now, suppose I venture to promise myself to cut my hair on Monday. Can I refuse to accept the offered promise? And once I accept it, can I release myself from the obligation? And how would that be different from simply changing my mind? Or let us say it is Tuesday now. So I am late. And I only have myself to blame. Should I apologize to myself? And if I do, should I forgive myself? How do I make myself trust myself again? Perhaps I have already damaged my relation to myself too severely and it is time to part ways. The last bit is patent nonsense. No one would be left after that divorce. The other verbs are, in fact, used reflexively in ordinary parlance. But whatever the proper interpretation of such talk might be – that much should be clear: the application of such concepts as forgiving someone or apologizing to someone presupposes the intelligibility of the question whether justice has been done to the other.[217]

This is a significant challenge. I am committed to the view that in order to count as having performed a speech act, a speaker must in some sense be "bound" by the constitutive norms of the speech act type they are performing. The question of exactly what relationship the speaker who engages in inner speech acts bears to herself—and especially what it could mean for her to "obligate" herself by norms in so speaking—will be addressed in Part IV. One of my concerns in that discussion will be to show that performing a speech act does not require intending to address an audience (much less the audience of oneself). The other will be to demonstrate that the notion of "obligating oneself" that is required to make sense of inner speech acts is perfectly comprehensible. However, this is of no use to Kantians like Korsgaard as the obligation in question is only conditional, not categorical. Speakers

[217] Haase (2014), p. 8.

who engage in silent conversations *must* conceive of themselves as subject to particular norms, but only given the fact that they want to count as performing particular speech acts.

Part IV: Inner Speech Acts

One purpose of the survey of empirical work on inner speech in Part III was to demonstrate that the distinction between dialogic and non-dialogic inner speech is an important one. It is hard to even imagine engaging in such inquiries without the resources to describe people as engaging in inner *conversation* as opposed to mere subvocal articulation. It is essential to our understanding of what inner speech is, as well as our ability to continue investigating it, that we recognize the possibility of *inner speech acts*—not just inner acts of speech. In speaking silently, we are able to *mean* what we say with various kinds of speech act force. We are able to assert, ask questions, conjecture, resolve, exclaim, suppose, wonder, swear, etc. Any theory describing what it is to perform a speech act that excludes this possibility is inadequate. Any theory that is able to describe the social, private, and inner levels of speech acts in a unified way is at a significant advantage. Throughout the next five sections I develop an approach to speech act theory that is able to do just that.

Section 1: Speaker Meaning without Communicative Intention

Up until this point I have been relying on a particular understanding of speech acts that may be described as a version of force conventionalism. According to force

conventionalism, in order for a speaker's utterance to count as the performance of a speech act it must be the case that (1) there exists a conventional practice corresponding to the speech act type in question with a unique set of constitutive norms,[218] (2) the speaker intends her utterance to have the force of that speech act type, and (3) the speaker's performance does not run afoul of Austin's category of MISFIRES pertaining to speech acts of that type (as discussed in the previous chapter).[219] This is the view advanced by the original founders of speech act theory, Austin and Searle, and extended by recent philosophers of language like Mitch Green and Dorit Bar-On.[220] However, there are other views of speech acts on offer, some of which will explicitly deny some of the claims that I will be making in this chapter. Some views stipulate that to perform a speech act one must achieve an effect on an audience.[221] Others require having reflexive communicative intentions.[222] On these and similar understandings of speech acts, it is by definition

[218] This is consistent with Ruth Millikan's biosemantic species of force conventionalism, according to which so-called 'natural conventions' are a type of convention. See Millikan (1998).

[219] Insincere speech acts are still speech acts. However, a speaker who intends to perform a speech act of a certain type but fails to properly invoke the convention in the first place has not performed the speech act. For example, by uttering "I hereby excommunicate you" I may intend to excommunicate you from the Roman Catholic Church. But if I am not the pope, my utterance fails to attain the speech act force of excommunication—it is a MISFIRE. In contrast, when I utter "I will meet you for lunch" intending this with the speech act force of a promise, I have successfully performed the act of promising *even if* I have no intention of following through.

[220] It is also generally friendly to Alston (2000)'s strategy for taxonomizing speech acts according to the responsibilities undertaken by speakers who perform them, though without committing to his accompanying theory of sentence meaning.

[221] For example, see A. P. Martinich's view that "audience uptake" is required for the successful performance of a speech act. See Martinich (1984), pp. 52–56. A more recent example of this view can be found in Kukla (2014).

[222] For example, see Bach and Harnish (1979) and Mark Jary's useful discussion of their view in his (2010), pp. 9–11. See also footnote 225 of this chapter below.

impossible to successfully perform speech acts when no audience is present (or at least, when we believe no audience is present).

Resistance to the idea of inner speech acts may feel natural. Intuitively, an audience *does* seem necessary for the performance of a speech act. For example, doesn't *asserting* necessarily involve *telling*? Even if my audience does not ultimately believe me, don't I need to at least *intend* to communicate in order for my utterance to rise to the level of assertion? Pre-theoretically, at least, it is tempting to say that *communication* is an essential goal of speaker meaning—speaking in such a way as to *mean* what one says with the force of a speech act rather than a mere act of speech. If there are such things as private and inner speech acts, then they would seem to be missing this key ingredient. Why not rather describe speakers engaged in such activities as *purporting* to perform speech acts, rather than actually performing them? Here is Mitch Green's response:

> A framed suspect might mean that she is innocent in saying, "I am innocent!", yet be fully aware that no one will believe her and perhaps, being realistic, not intending to convince anyone. She might not even intend her interrogators to believe that she believes she is innocent, since she might know that they are certain she is lying. Or, gazing into my newborn daughter's eyes I might say, "All things valuable are difficult as they are rare," meaning what I say, without having the slightest intention to produce beliefs or other attitudes in her or in anyone else. Again, in the film Sleeper, Woody Allen's character Miles Monroe comes across, while exploring alone, a genetically modified chicken the size of a small house. Miles remarks, "That's a big chicken." In saying this he does not seem to be intending to produce an effect on anyone, himself included.

Cases like these lead Green to offer an account of speaker meaning as "making manifest." To make an intention manifest is to intend that it be made accessible, whether

or not it is in fact discerned by anyone.[223] According to Green, for an utterance to rise to the level of speech act it is not necessary that the speaker intend to communicate—it is not even necessary that there *be* an audience capable of understanding what is said. All that is necessary (in addition to avoiding MISFIRES) is that a speaker *intend* for her utterance to have the right speech act force, and intend that *that* intention be made manifest—accessible even if not actually discerned.

To make this idea vivid, it will be helpful to consider another example. Imagine three people, each of whom devotedly keeps a diary: Jericho, Lala, and Gabriel. Jericho's diary is fairly ordinary, but he intends that no one ever read it. Still, *were* someone to read his diary, they would easily understand it. Jericho knows this, which is why he keeps it hidden. Lala's diary is similar to Jericho's, except that she intends to give it to her children one day. Lala wants her descendants to be able to learn about her life in her own words—and certainly, if they do read it one day it will be perfectly comprehensible. Gabriel's diary is very different. Instead of writing words, Gabriel has chosen a diary form that involves using random shapes and patterns to express the feelings and events that she experienced that day. Were someone to happen upon Gabriel's diary they would not understand any of it. Gabriel herself usually forgets what she meant by some particular doodle or other. It makes little difference to her whether anyone looks through the pages, for they will be incomprehensible.

What Green's theory suggests is that there is far more in common between Jericho and Lala's diary writing than either of them shares with Gabriel's. Both Jericho and Lala mean what they write with particular speech act force, and make their intentions manifest

[223] Green (2007), p. 65.

in a way that is accessible—whether or not their diaries are ever actually read. The fact that this act is a communicative one for Lala (she intends for her words to have an audience) while it is decidedly not for Jericho makes no difference to whether or not they have performed speech acts. It is precisely *because* Jericho's diary makes his intentions accessible that he keeps it hidden, in order to prevent his act from being a communicative one. In contrast, despite the fact that Gabriel's sketches are in some sense expressive of her inner states, they fall short of speaker meaning and speech act-hood.

The example of the diaries allows us to clarify an aspect of Green's account of speaker meaning. Remember that making manifest involves "making accessible," whether or not what one says is comprehended by an audience. This might seem to rule out the possibility of speaker meaning in inner speech, as the contents of our own minds are paradigm examples of that which is publically *inaccessible*. However, this challenge only arises if we conflate two different senses of accessibility. Our inner speech is inaccessible in the same way that Jericho's locked away diary is inaccessible: no eavesdropper or nosy reader can *actually* get to them.[224] However, *were* someone able to find Jericho's diary, or listen in on our inner speech, they *could* perceive the relevant intentions to mean what is said with the force of a speech act. This is the sense of accessibility that ensures Jericho's diary and our inner speech are capable of rising to the level of speaker meaning, while Gabriel's spontaneous shapes and lines do not—even if she puts her diary on public display as an art installation.

[224] Perhaps it could be replied that *God is always listening*. Even if this is true, speakers (even religious ones) often at least *believe* that no one is listening. In such cases their utterances—whether audible, silent, or written—may still rise to the level of speaker meaning.

The diary case, Green's examples, and the empirical work on inner speech canvased above together make a compelling case for the plausibility of performing genuine speech acts when no audience is present—speaker meaning without the intention to communicate. To be sure, communication is an important goal of the most paradigmatic examples of speech acts. But it is not *necessary*. Recognizing this removes one potential barrier to affirming the existence of inner speech acts. The framework of force conventionalism—together with Green's definition of speaker meaning—allows us the opportunity to go about constructing a unified understanding of speech acts in both communicative and non-communicative contexts. Other views in speech act theory, such as those that require producing an effect on an audience or even reflexive communicative intentions, rule out the possibility of solitary speech acts from the start. This puts such views at a significant explanatory disadvantage.[225]

A more interesting and instructive point of disagreement comes from recent work by Mikhail Kissine. Kissine agrees that what it takes for an utterance to rise to the level of speech act is for the speaker to intend her utterance with the force of that speech act. That is to say that he accepts the general framework of force conventionalism. However, Kissine doubts that speakers who engage in private talk or inner speech *can* intend their utterances

[225] Unless, of course, one goes in for a literal interpretation of Arendt and allows (like Korsgaard and Darwall) for the possibility that one may address oneself in exactly the same way that one addresses another. This is the strategy pursued by Schiffer (1972), Strawson (1970), Bennett (1976), and Avramides (1989). I have set such a proposal aside in the interest of determining whether force conventionalism is consistent with truly audience-free speech act performance. This question is important regardless of whether inner speech involves self-address because there are other examples of unaddressed speech (noted above) that we intuitively think ought to count as speech acts. Such views also tend to require highly complex iterations of Gricean reflexive communicative intentions for speakers to count as exhibiting speaker meaning and performing speech acts. For a critical discussion of such views see Green (2007) pp. 61–69.

to constitute speech acts. His assumption seems to be that part of what it is to *mean* one's utterance with a certain speech act force requires *addressing* it to an audience.[226] Kissine thinks that even in cases where one's private and inner speech seems to be addressed towards oneself, it still cannot rise to the level of genuine speech act. For support he appeals to Erving Goffman's observation that "in our society at least, self-talk is not dignified as constituting an official claim upon its sender-recipient....There are no circumstances in which we can say I'm sorry, I can't come right now; I'm busy talking to myself."[227]

To see why Goffman's point raises problems for the possibility of private and inner speech acts, notice that force conventionalism implies that part of what it is to perform a speech act of a given type is to be in some sense "obligated" by the norms constitutive of that practice. As discussed in the previous chapter, this explains how it is that speakers are able to *express* the mental states required by the sincerity conditions of the speech act type they are performing. In the context of a social speech act, the situation is relatively straightforward. Provided that communication is successful, the speaker is bound to her speech act's norms *by her audience*. That is to say, by performing a speech act of a certain type, the speaker sticks her neck out by licensing her audience to criticize her in the event that it is discovered that she is insincere in what she says. In the context of private speech and inner speech, it is much less clear what the source of such a binding or "obligating" force could be (especially given that the purpose of such acts is not *communication* in the ordinary sense).

---

[226] Kissine (2013), p. 23.
[227] Goffman (1978), p. 788.

Kissine's skepticism about inner speech acts echoes Haase's observations raised above. As Haase says, "the talk of obligating risks becoming empty, if the one who obligates and the one who gets obligated are identical…if the one who has the power to release is the same as the one who is under the obligation, then there is no obligation."[228] While I have made sense of the idea of speaker meaning without addressing an audience, this really just pushes the problem back a step. If the intelligibility of performing a speech act requires being obligated by the constitutive rules of the speech act type, then according to Kissine and Haase the necessity of an audience returns. The coherence of inner speech acts is still in doubt.

Let's take stock: The only theory of speech acts on offer that does not rule out the possibility of private or inner speech acts by fiat is something like Force Conventionalism. Other theories of speech acts rule out the possibility of private or inner speech from the get go, by including reference to an audience in the definition of what it is to perform a speech act in the first place (intending to produce an effect or possessing a reflexive communicative intention). However, in order for FC to make sense of the phenomena of private and inner speech acts, it must be possible for speakers to in some sense "bind" or "obligate" themselves by the norms of the speech acts they are performing. Kissine agrees that force conventionalism is true, but like Haase is skeptical about the possibility of a speaker binding themselves in the absence of an audience. This leads him to deny the possibility of private or inner speech acts. Fortunately, one philosopher's *modus tollens* is another's *modus ponens*. As discussed above, we have good reason to endorse the possibility of private and inner speech acts—a "dialogic" way of speaking that is distinct

[228] Haase (2014), p. 8.

from non-dialogic private and inner *acts of speech*. Because the only way to make sense of this is to adopt a theory like Force Conventionalism, I argue that we have strong reasons to believe that *it is* possible for a speaker to—in the relevant sense—"bind" themselves by the constitutive norms of the speech acts they are performing.

Having strong *reason to believe* that it is possible for a speaker to bind herself by norms is not the same thing as fully *understanding how* the mechanism works. Hence I owe more of an explanation, and in fact I do believe that there are a few things that we can say about it. For one thing, we need not think that the social dimension is ever absent entirely. If the Vygotskians are right, then social speech is developmentally prior to private and inner speech. Perhaps it is incoherent to imagine that the speaker herself could be the absolute *origin* of the binding force of her speech act norms. If so, then we also have a strong reason to favor the Vygotskian picture. To that end, I will now turn to an overview of how self-binding might work, and how it is dependent on social speech.

Section 2: Viewing Oneself as Constrained by a Norm

In what follows I will be relying on a subtle but important distinction. This is the difference between actually *being bound* by a norm, and *taking oneself* to be bound by a norm. Often these two will come together. I take myself to be bound by the moral norm that prohibits needlessly taking innocent life. Plausibly, I am actually bound by this norm as well. Similarly, I might take myself to be bound by the norm of etiquette that recommends waiting until all at the table are served before beginning to eat. In many contexts, this norm also actually applies to me. Another example: I take myself to be bound by the speech act norm that requires belief in what I assert. As a matter of fact, I am

probably bound by this norm as well. In order to count as having performed an assertion, among other things I must count as a rule breaker if I do not believe what I say.

Sometimes, however, this happy situation does not obtain and the norms that actually apply to me may not be the same as the ones I take myself to be bound by. I may believe that I am on a "first name basis" with my colleagues when actually I am not. In such a situation, a norm of etiquette applies to me even though I do not see myself as bound by it. Or perhaps I take myself to be bound by a moral norm that forbids consuming caffeine when in reality no such moral norm exists. Finally, I might take myself to be bound by a speech act norm that requires absolute certainty in every assertion I make. Probably I am not actually bound by such a norm.

I would like to point out two things with respect to this distinction. Firstly, even if the most uncompromising nihilist about normative properties ends up being right,[229] it would still be the case that most of the people in the world *take themselves* to be bound by norms. In perhaps a less farfetched scenario, it could be true that norms within a certain domain do not exist but that nevertheless many people take themselves to be bound by those norms. For example, one could think that there really are no such things as "norms of etiquette" or "norms of morality." But one would be hard pressed to deny that many people *take themselves* to be bound by such norms.

The second thing I would like to point out about this distinction is that, for the purposes of explaining deliberation and action, what matters most are the norms a person *takes themselves* to be bound by. On the one hand this means that when I am deciding what

[229] We must set aside for the moment complications that arise from trying to determine what "being right" could mean in such a situation.

to do, or what I believe, the fact that I take myself to be bound by some relevant norm can have a causal influence on the course of deliberation or action I end up pursuing. For example, my believing that drinking caffeine is wrong is a fact that could factor into explanations about why I deliberate and act in the ways that I do—regardless of whether such a moral norm exists or not. On the other hand, if an agent does not take themselves to be bound by a norm, then the fact of the norm's actual existence will make little difference. Suppose there exists an aesthetic norm that prohibits wearing both black and brown articles of clothing at the same time. If I do not believe such a norm exists, or at least do not take such a norm to apply to me, then the fact of its *actual* existence will make no difference to how I deliberate or behave.

There may be views about the nature of normativity according to which the mere existence of a norm can make a difference to the deliberations or actions of an agent, regardless of whether the agent herself is aware of that norm's existence. At the very least, it might be said that a norm's existence could play some causal role in an agent's coming to believe or know about it. For my purposes here, I believe I can remain neutral with respect to these possibilities. What I will maintain—and what I do not think that anyone would deny—is the initial observation running in the other direction: that taking oneself to be bound by a norm can make a difference to an agent's deliberation and action, regardless of whether the agent is actually so bound. It is this latter claim that will prove most important in what follows. To sum up our observations: (1) whether or not a norm actually exists and/or applies to an agent, that agent can still *take themselves* to be bound by it; and (2) whether or not a norm exists and/or applies, if a person *takes themselves* to be bound by it then that fact can induce them to heed it.

Now let us return to speech acts. According to Force Conventionalism, for an utterance to rise to the level of speech act a speaker must intend to put it forward with the force of that speech act type. Because speech acts are defined by their constitutive norms, to intend for one's utterance to count as a speech act of a certain type is (at least partially) to intend for oneself to be bound by the norms constitutive of speech acts of that type. For example, part of what it is for me to intend that my utterance has the force of an *assertion* is to intend that I myself count as deserving of criticism if I do not believe what I said. Because what matters is a speaker's *intending to be bound* by these speech act norms, I submit that their utterances can rise to the level of speech act as long as they *take themselves* to be constrained by these norms—regardless of whether they are *actually* bound by these norms. As it turns out, the crucial idea for the maintaining the coherence of inner speech acts is the concept of *intending to bind oneself by a norm*. Before we can proceed, however, this needs to be distinguished from merely *deciding to follow a norm*.

What does it mean to describe someone as "intending to bind themselves" by a norm or rule? Specifically, what is the difference between deciding to be *bound* by a rule and merely deciding to follow a rule?[230] If there is no difference, then force conventionalism risks triviality. Consider a potentially analogous case. Inspired by Rodin's sculpture, suppose I decide to rest my chin upon my hand. In so doing I realize that my action will only be able to count as being of the type intended if my chin actually makes contact with my hand. We could describe the situation in this way, as a parody of my version of Force Conventionalism: chin-on-hand-resting is an action that possesses as one

[230] This challenge is based on Raymond Geuss' criticism of Korsgaard in *The Sources of Normativity*. See Geuss (1996) p. 191.

of its constitutive norms the rule that the rester's chin must be in contact with the rester's hand. By deciding to perform the action of resting one's hand on one's chin, a rester chooses to "bind" themselves by the norms constitutive of actions of that type. If choosing to be "bound" by a rule is no different from choosing to follow a rule, then we can redescribe any action as having this structure. There is no more to "binding" oneself in this sense than merely recognizing that *in order to count as having done something, one must actually do it.*

The easiest way to see that a speaker's binding of herself by the constitutive norms of the speech act she performs amounts to *more* than merely deciding to follow a rule is to take a closer look at infelicitous speech acts like lying. In paradigmatic cases of lying, a speaker asserts something that they consciously believe to be false. This is an instance of a speaker choosing to perform an action that is constituted by certain norms, and at the same time intending *not* to follow those norms. According to force conventionalism, liars *bind* themselves by the norms constitutive of assertion without deciding to follow those norms—in fact they *intend* to violate them. To bind oneself by a norm is *not* to decide to follow the norm. Rather to "bind" oneself is to choose for a norm to apply to oneself, to choose to fall under its jurisdiction. It is very important to liars that the norms of assertion apply to them—that they count as criticizable in the event that their deception is found out. This is not to say that most liars *want to be criticized*, because most liars do not wish to be found out. But they do want to count *as deserving of criticism.* It is crucial to pulling off their deception that what they say is believed, and hence it is essential that their action counts as being one of asserting. One cannot successfully violate a rule that does not apply. If a speaker does not bind himself by the norms of assertion, then he cannot tell a lie.

Usually speakers who perform speech acts intend *both* to bind themselves by the constitutive norms of the type of speech act they perform *and* to actually follow those norms. That is because most speakers mean what they say in good faith; they are sincere. However, the two intentions may be distinguished from one another. Binding oneself by speech act norms involves more than the recognition that counting as having done something requires actually doing it. This is because, in the case of certain conventional actions like speech acts, one may perform them felicitously or infelicitously. I cannot rest my hand on my chin infelicitously: either I rest it on my chin or I do not, there is no room for an "improper" hand-on-chin resting. However, I can play chess infelicitously: I can cheat. A cheater only counts as cheating because they have chosen to play a game that is constituted by rules and at the same time have chosen to break those same rules. In the same way, a speaker may perform a speech act infelicitously—by accident or by design.

To decide to perform a speech act of a particular type is to choose for the norms that govern an institution to apply to oneself. As long as the speech act is performed successfully, those norms do in fact apply. But the fact that a person has chosen to perform a speech act, and the fact that the norms of that speech act consequently apply to them, are together still insufficient to settle the question of whether the speaker has chosen to follow the norms of the speech act. A further fact must be added to the story: to what degree is the speaker trying to be sincere? Speech acts, like board games and many other cultural artifacts are socially defined conventional activities. As such they are constituted both by conditions for counting as having occurred at all, and also by conditions for having come off happily. In order to count as having pulled off such a conventional act, one must view oneself as assessable by its felicity conditions. In order for such a pulling-off to be happy,

it must fare well by the lights of those conditions. In the next section I will put this story to the test in one of the most difficult examples for the inner speech act theory.

Section 3: Can I Make a Promise to Myself?

To assess the coherence of my proposed understanding of self-binding, let us consider one of the most difficult examples for my view: promising. According to force conventionalism, for a speaker's utterance of "I will finish the paper by Monday" to count as a *promise* it must be the case that the speaker intends for her utterance to have the force of a promise. This means that the speaker must intend to be bound by the norms constitutive of the speech act of promising. In other words, the speaker must intend to count as justly assessable by promising's felicity conditions. Among other things, this means that the speaker has decided to be deserving of criticism if it turns out that she had no intention of doing what she said (an example of Austin's category of INSINCERITIES). She has also chosen to count as deserving of criticism in the event that she fails to accomplish what she promised (an example of Austin's category of BREACHES).

Another important feature of promises (as pointed out by Haase above) is that the promis*or* gives the promis*ee* the power to release them from the promise at will. In the case of social promises, the binding quality of this relationship is clear. However, one might worry that it is incoherent in the int*ra*-personal case. If a speaker makes a promise to themselves, they are both promisor and promisee. But how can a promise be binding if the promisor has the power to release themselves at will? Doesn't this undermine the whole point of promising? If a promise is not binding, it cannot count as a promise at all (even according to Force Conventionalism).

Let us consider a fictional story.

> Julian cares deeply about his own physical fitness and health. He becomes convinced that the best thing for him would be to begin cutting refined sugars out of his diet. He finds himself in a difficult situation, as he truly loves drinking soda—diet soft drinks just won't do. Nevertheless, he values his long-term health over the temporary pleasure that consuming refined sugars brings him. He makes a promise to himself *sotto voce*: *"For the next three months, I promise that I will not drink any sodas containing sugar."*
>
> The first month goes well, Julian feels great. The second month the cravings begin. With time and perseverance his body eventually becomes used to the feeling of life without all that sugar. Consequently, the benefits of giving up soda start to become less noticeable. Every now and then Julian thinks to himself, *"I'm doing so well, perhaps a soda here and there wouldn't be so bad…"* This always prompts another thought to follow soon after, *"but I promised—just a few more days to go."* A week before the three months are up, Julian is out with some friends. He hasn't seen them a while, and they have brought along a case of Guarana Antarctica—a soda they all discovered together on a high school trip to Brazil. Julian, caught up in the moment, forgets his promise and has a sip of the delicious drink.
>
> As soon as the sugar hits his palate, the memory of his promise comes rushing back to Julian. He is overcome by guilt for his failure. *"I'm sure they would have saved me some of this Guarana if I had asked…Why couldn't I just have waited a few more days?"* he berates himself, regret washing over his face. Julian's friends, noticing the change in his countenance, quickly ask him what is the matter and Julian explains his failure. His friend Tasha tries to console him, "C'mon Jules, you didn't do anything wrong."
>
> "Maybe," Julian replies. "But I still feel bad."

What this story illustrates is that, though we can make promises to ourselves, breaking such promises is not typically described as moral wrongdoing. Nevertheless, it is still felt to be a failing—Julian's guilt is evidence that he took himself to be bound by the norms constitutive of promising. Because he is both promisor and promisee, it may be too strong

to say that a moral obligation has been violated. However, it still feels like a *broken promise*.

In contrast, consider an alternative ending to the story in which Julian explicitly releases himself from the promise before he drinks the soda. Perhaps he deliberates carefully before he acts and decides that the value of the shared experience with his friends outweighs the importance of sticking out the last few days of the promised three months. In such a case Julian *releases himself* from the promise. If he feels bad afterward, it will be because he did not end up accomplishing the goal he set himself. But this will not be an "all things considered" kind of guilt like he felt in the first story. Rather it would be the bittersweet feeling of loss when one value outweighs another. In a situation in which Julian releases himself from the promise before drinking the soda, the emotion he feels is not *remorse* for making the choice he did. Rather it would be *regret* that circumstances obtained which forced him to choose between two values: sharing an experience with friends vs. achieving a personal goal. Unlike the original story, he would not feel the specific sensation of *having broken a promise*, only that of giving up one value in favor of another. It is the difference between guilt and disappointment, though of course in this case (one hopes!) the emotions are not terribly intense. It is just a soda, after all.

What I believe this shows is that there is a distinction between the *moral* and the *pragmatic* dimensions of the norms that can be created by a speaker who makes a promise. On the one hand it does seem plausible that a moral obligation is created when a speaker successfully makes a promise that is accepted by an audience. However, if the promisor and promisee are the same person then perhaps, following Haase, it is incoherent to suppose that such a categorical obligation could obtain. If Haase is right, then to say that I

am under a moral obligation that I may release myself from at will is to say that I am under no moral obligation at all. This seems reasonable to me. However, pragmatic norms are not meant to be categorical in this way. To count as subject to the norms constitutive of speech acts of a particular type, all a speaker must do is intend their performance to be an action of that type (regardless of whether they also intend to follow the relevant norms). Perhaps a speaker cannot bind themselves to a moral obligation by making a promise to themselves, but they can make a resolution.[231]

What is the difference between merely deciding to do something, and *resolving* to do it? What extra weight does "promising to oneself" add to our intention to do something, if it is incapable of establishing a moral obligation? This is an interesting question, worthy of further investigation. If that which we promise to do is already something about which we care a great deal, it may feel inappropriate to criticize a failure to follow through for its *infelicity*. Presumably, the grief at missing the mark with respect to something that we greatly value renders mute self-criticism for the additional failure of satisfying the norms of the speech act we happened to have antecedently performed. Why then do we make promises to ourselves at all? Why not just concentrate on the value that we already placed on the desired goal?

The first thing to point out is that the disinclination to blame ourselves for infelicity does not on its own cast doubt on the fact that we have fallen short of the norms of the speech act we have performed. When a single action runs afoul of two values one of which

[231] One surprising result of this discussion is that views according to which the constitutive norms of speech acts *just are* moral norms may end up being inconsistent with the possibility of performing speech acts in the absence of an audience. I take this to be a significant problem. For a recent example of such a view, see Cuneo (2014).

is far greater than the other, it may feel fatuous to criticize the actor for his failure with respect to the trivial value—especially as one stands face to face with the loss of a great value. It may be easier to feel the weight of the norms of the speech act of promising if we consider cases in which a speaker promises to do something about which they do not care very deeply. However, if the object is too trivial, it will be easy for the speaker to change her mind and hence difficult to count as breaking the promise. We seem to be faced with a dilemma: if what we promise ourselves is something that we have a great independent desire to do, then the weight of the promise seems trivial by comparison; but if it is something that we do not care much about, then it will be easy for us to change our minds. It is thus hard to see what point there is to making promises to ourselves, even if it is something that we are technically capable of doing.

I will offer two suggestions in an attempt to quiet the force of this dilemma before moving on. The first is that we can make promises at times when we are afraid that our great desire for the object may not be enough to ensure fidelity. In such cases, when there are contradictory impulses inside of us, making our commitment explicit can help to tip the scales. This seems to have been Julian's situation. Pursuing a healthier diet is something that he cares deeply about, but he also recognizes that he will have difficulty always sticking to it. By giving specific shape to a nebulous goal, Julian's promise can help him to fortify his will.

The second suggestion is that we may often make promises simply to express our intentions—to make them manifest—even when we do not foresee that fulfilling them will be terribly arduous or contrary to impulse. Recall that one of the consequences of a speech act type's requiring a speaker to possess a particular state of mind is that performing speech

acts of that type enable a speaker to *stick her neck out* and thereby express that state of mind. This is how making a promise enables a speaker to express an intention. And the expressive quality of promises can help to explain more than just a possible motivation for making a promise to oneself. It can also help to explain what we are up to when we make promises to audiences incapable of holding us to them (or even of understanding what we are saying at all). When a parent promises to an infant child, when an owner promises to a pet, when a user promises to an appliance—in each of these cases the speaker's goal is to making his or her intentions manifest.[232] This casts more light on the first suggestion: part of what we remind ourselves when we recall a self-promise to strengthen our will in a time of weakness *just is* the memory of a moment when our intention was vividly expressed. I will return to this mechanism briefly during the discussion of normative speech acts at the end of this chapter.

The argument of this chapter does not stand or fall on its account of promises. I wished to examine them because promising is, in one sense, the most challenging speech act to imagine successfully performing in a solitary context. Even if one is not convinced about the coherence of self-promises in particular, the discussion has led to an important more general point. What determines the speech act type a speaker has performed is not the goal which the speaker has in mind by performing it, but the felicity conditions to which she subjects herself. Certainly, it may be true that a speech act of a given type was originally

---

[232] Or take a more unusual case. When the villain promises to destroy all that the hero holds dear it is not as though a moral obligation is created. After all, the villain is promising to do something *that the hero doesn't want.* Does the villain fail to make a promise at all? I suggest that the villain is making a promise to himself, and is threatening the hero by informing her of this fact. Even though a moral obligation is not created, by promising the villain is able to express his evil intentions—and this precisely in virtue of the felicity conditions of the speech act he performs.

designed to accomplish a particular goal or narrow range of goals. That would explain how it came to have the felicity conditions it does.

For example, perhaps the institution of assertion was originally designed only for the goal of communicating reliable testimony from one speaker to another. This might explain why assertion developed as a convention to have the particular rules that it does. However, now that we *have* the convention of asserting available to us, we *may* make use of it to accomplish a variety of different goals. In the same way, perhaps the game of chess originally arose as a way for two individuals to compete against each other—each with the goal of winning against the other. But now that we have the game of chess, we can play it with a variety of different goals in mind. I may play a game of chess with a small child with the goal of losing—just to have fun, to teach them the game, or to cheer them up. I can play chess against a skilled opponent with the goal of achieving a stalemate—as a personal challenge, or to cause them annoyance, or to work towards a better understanding of the game. In each case it *is* essential to playing chess that I view myself as constrained by the rules, but it is *not* essential that I restrict myself to a particular narrow range of goals.

The general lesson is that the very same type of speech act may be performed for a variety of reasons, while at the same time *the speech act's having exactly the felicity conditions it does* will still be essential to its accomplishing each of those various purposes. I may assert "it is raining again" with the goal of answering your question about the weather, or of defending my desire not to take a walk, or of explaining why I arrived home soaking wet, or merely to make awkward small talk with my barber. In each case making reference to the constitutive norms of assertion is essential to understanding *what I have done*. I couldn't accomplish what I set out to accomplish by saying what I did if my

audience did not take me to believe what I was saying. But there may be many different reasons *why I have done it*. Elizabeth Anscombe makes this same point about the speech act type of *commanding*:

> A command will be a description of some future action, addressed to the prospective agent, and cast in a form whose point in the language is to make the person do what is described. I say that this is its point in the language, rather than that it is the purpose of the speaker, partly because the speaker might of course give an order with some purpose quite other than that it should be executed (e.g. so that it should *not* be executed), without detriment to its being an order.[233]

The same is true of other cases that Haase emphasizes, like asking questions and making promises. It could be that my goals in making a promise to myself and making a promise to another are very different. And it is certainly true that there are many goals of promising that I could not intelligibly have in view when I do it in the privacy of my own thoughts—at least in the non-pathological case. I cannot coherently intend to place myself under a categorical moral obligation, for example. But I can intend to establish a personal resolution, and it is essential to accomplishing this goal that I perform the speech act that I do—one that includes in its felicity conditions that I am criticizable both if I do not intend to do what I say, and also if I ultimately fail to follow through (without first releasing myself). This is why it is important that we can see the difference between the case in which Julian releases himself from the promise and the case in which he doesn't. In neither case does it seem right to describe him as having failed a moral obligation, but in one case it does seem right to describe him as having *broken a promise*. When we make promises to

[233] Anscombe (1957), p. 3.

ourselves, the difference between *weakness of will* and *changing one's mind* is a significant one.

Such a use of promising exists in the interpersonal context as well. Imagine I say to an old friend "let us promise to meet here again in ten years' time." If the promise is made, it is a resolution by the two of us to do something *together*. After five years my friend could write me a letter, "I have changed my mind—you don't have to meet me after all. I will not hold you to it." Either of us can release the other from a promised joint action and, by so doing, release ourselves. This is much different from a case in which no release is given and instead the promise is broken. Having the power to release oneself does not preclude the possibility of weakness of will. Perhaps our collective promise could not have had the goal of creating a morally binding obligation, but that does not mean it was not a promise. It just means that the norm we have bound ourselves by is a weaker one, conditional for its existence on the fact that neither of us changes our minds.[234]

Consider also what Haase says about the speech act of questioning:

> There is *a* sense in which one can ask oneself a question and answer it. It does not follow that this is the *same* sense in which one can ask another and get a response. On the face of it, the former is just another way of saying that one wonders, say, whether *p* and then settles the question. And that formulation does not invoke the picture of an interaction with oneself. In any case, the application of the concept *asking someone* presupposes the

[234] Consider the analogy of cooperative board games—games in which all players work together to defeat the game itself, with no actual player in the role of the opposition (see for example games like *Arkham Horror* and *Pandemic*). There is a big difference between a case in which the players discover at the end of the game that they have been breaking a rule, versus a case in which the players collectively decide to modify the rules. In such cases the difference between weakness of will and changing one's (collective) mind may determine whether or not the group conceives of themselves as having won the game—as having accomplished the goal they set themselves.

> intelligibility of the scenario in which the one who answers *informs* the one who asked—by telling her, by teaching her or by giving testimony to her.[235]

For Haase, the "speech acts" we perform silently are but shadows of the speech acts we perform to others. But this does not follow. It is true that the reasons *why* we ask questions of others and ask questions silently when we are by ourselves are sometimes different. And it could be that there are some reasons for asking questions that simply do not arise in monologue—setting aside pathological cases. What Haase says is plausible: for me to ask a question *with the goal of obtaining information*, I must suppose that my audience knows something that I do not. When we ask questions silently in inner speech, we are not typically requesting that we inform ourselves.

But all Haase's point shows is that there are *some* reasons for asking questions that presuppose an audience. If my goal in asking a question is to request information, or mislead someone, or to teach (perhaps according to the Socratic method), these are not ordinarily things I would be aiming at in a solitary context. But there *are* reasons for asking questions that *do* make sense even when I am by myself. I might be identifying a practical problem ("where should I eat for lunch?"), or raising a topic as the object of investigation ("is there a coherent analysis of inner speech acts?"), or expressing amazement ("how did the magician do that?"). And these same goals make sense in the interpersonal case as well. We may be trying to solve a practical problem *together*, collectively engage in an intellectual inquiry, or express our puzzlement to one another. Requesting information is not the only reason we ask questions. To say that a person is *merely* "wondering whether" or "posing a question" is not to say that they fall short of performing the speech act of

[235] Haase (2014), p. 4.

questioning, but to begin to identify what their goal was by performing it—whether silently or aloud. Charles Fernyhough describes an empirical investigation into one of the possible functions of self-questioning:

> It may be that something about the linguistic act of posing a question to yourself can make your intentions about what you are planning usefully clear….This idea was tested in a study by the psychologist Ibrahim Senay and his colleagues at the University of Illinois at Urbana-Champaign. They gave participants a task involving solving anagrams, but asked them to prepare for it in silence, either by asking themselves questions about what they were about to do, or simply by making statements about it. When they were instructed to question themselves silently, the volunteers solved more anagrams than when they merely declared their plans to themselves in inner speech. The researchers concluded that quizzing yourself in self-talk can push you beyond what you might otherwise achieve if your inner speech is full of bald statements of intention.[236]

Ultimately Haase shares in a misconception raised near the beginning of this chapter. His emphasis on communication—understood narrowly as *the mere exchange of information*—as the only goal of social conversations is what fuels his unwarranted skepticism about a truly dialogic mode of inner speech. Once we expand our theory of speech acts to better capture the rich diversity of form and purpose that social speech can take, we see that room for phenomena like inner speech acts has already been made.

When performing speech acts, it is significant that we are able to distinguish between insincerity, duplicity, weakness of will, and changing one's mind. If all there were to "binding oneself" was just deciding to follow a rule, none of the various forms of infelicity would be possible—there would only be changing one's mind. This is just as true in the *intra*-personal case as it is in the *inter*-personal case. It may be that there are certain

[236] Fernyhough (2016), p. 105.

reasons *for* performing speech acts that do not arise when we are alone. However, no matter our reason for performing a speech act of a particular type, it is still necessary that we conceive of ourselves as subject to the constitutive norms of speech acts of that type in order for what we are doing to be intelligible to ourselves as being an action *of that type*. When I assert something silently, I take myself to be bound by the same constitutive norms of assertion that would constrain the social version of my speech act. This is what marks my silent utterance out as a *speech act* rather than a mere *act of speech*, and what marks it out as a speech act of assertion rather than of some other type.

Section 4: Bringing Inner Speech Acts into View

Is it possible for speakers to perform silent speech acts *insincerely*? What could it mean to say that I have made a "false promise" to myself, or that I have "lied" to myself? I believe that there is one sense in which it is extremely difficult to engage in infelicitous inner speech, and another sense in which it is very likely that we slip into infelicity—at least from time to time. It does seem difficult, if not impossible, to intend for one's silent speech act to have a particular speech act force while at the same time being consciously aware that you do not satisfy the norms of that speech act type. For example, it hard to imagine silently uttering something with the full force of an assertion that one consciously does not believe. A sentence expressing a proposition the thinker does not believe could be silently uttered with the force of *supposition for the sake of argument*, or with the force of an actor as though on stage portraying a character, or imagined as being asserted by someone else. But it is not easy to conceive what it would be like to silently perform a genuine assertion that one does not believe, while at the same being fully aware of one's

insincerity. I do not wish to completely rule out the possibility, only to recognize that the experience would be a strange one—difficult even if possible.

At the same time, just as in public speech, our silent speech acts may be infelicitous because of failures of self-awareness. I may promise something while lacking the relevant intention to follow through, assert something without the appropriate belief, apologize for something without feeling sincere remorse, etc. all because—at the moment of utterance—I lacked the self-awareness to see that I did not possess the mental state in question. In fact, it would be very difficult to go one's whole life without ever unintentionally falling short of full sincerity for this reason, even in the privacy of one's own thoughts. Often, however, when our speech acts do happen to fall short of wholeheartedness, the very act of speaking will make us conscious of our insincerity. If I do not believe what I have silently said, my assertion can *ring hollow*—I can be made suddenly aware of my own lack of belief in what I have said (or even of my own positive belief to the contrary). There may also be times when our minds have not yet been fully made up with respect to the subject about which we speak. In such cases the speech acts we perform can cause us to *come to take a stand* one way or the other by confronting us with the sincerity conditions of our utterance.

To illustrate these phenomena, imagine that there were times in the past when the character Julian from our story above asserted silently to himself, "*sodas don't really have that much sugar*." Given what we know about Julian, we can be confident that over time his assertions began to *ring hollow*. At first Julian really does believe what he says. Gradually, however, he comes to doubt his justification. The act of asserting creates a psychological pressure to either come up with new justification or revise his beliefs. This is because the act of silently asserting made salient to his occurrent thought the sincerity

conditions constitutive of his action. As he researches the sugar content of soda, Julian's beliefs begin to change. The last time he asserts "*sodas don't really have that much sugar*" he doesn't believe it at all. The dissonance he feels may not be guilt of *having inadvertently told a lie*, but it is still the sensation of *having said something that one comes to realize wasn't fully sincere*. Julian's speech act echoes emptily, a sensation which might cause him to quickly follow it up with a correction (another assertion): "*I don't really believe that…*"

We can imagine a similar kind of phenomenon occurring in the use of irony. Suppose that Julian enjoys sarcasm. This is part of his personality, and it shows through in his inner speech as well. He used to silently say things like "*oh sure, soda is just so bad for you*" with an inner sneer. The sarcastic speech act plausibly has constitutive norms too—something to the effect that you ought to believe the opposite of what you have said.[237] As Julian's thinking about the harmful effects of sugar changes over time, these sarcastic remarks begin to ring hollow. Eventually he is struck by his own insincerity: "*Wait—I think I really do believe that.*"

The cost of such insincerity is not, as in the public case, a risk of criticism from an audience. Rather, the cost is a deliberative one. I cannot build upon my assertion in the course of deciding what to do or determining what I believe. I fail to add to my internal conversation's stock of consciously accessible "common ground." The cost of violating the norms of your inner speech acts is *an inability to put that speech act to use in accomplishing whatever it was you were trying to do*. Even though our goals in performing

[237] I am not aware of existing theories of the nature of sarcastic or ironic speech that analyze it as its own distinct type of speech act. However, it seems like an obvious approach to take. For a fascinating (though troubling) look at one particular subculture for whom the norms of sarcasm, irony, and trolling are very important—and in which being a true believer can be a dangerous liability—see Penny (2016).

speech acts in inner speech are sometimes different from our goals in performing speech acts publically, it is still essential to accomplishing our goals that the speech acts have the norms they do, and often it is important that we satisfy them as well. After all, that is why we were speaking with the force of *that* speech act type, rather than another.

Let's summarize the argument so far: in many cases it is clearly possible for us to view ourselves as bound by the norms constitutive of the speech acts we perform silently to ourselves. *Taking oneself to be bound* is sufficient for *intending* to perform speech acts of the respective type. According to Force Conventionalism, acting with this intention is enough to fully count as having performed a speech act. Speech acts are acts that can be performed simply by *meaning* what one says with the right kind of force, and to have performed such an act is to be bound by the norms constitutive of acts of that type. Therefore, it appears that in this particular case our distinction has collapsed: *taking oneself to be bound* by the norms constitutive of a speech act, because it is sufficient for having *actually performed* the speech act, is in turn sufficient for *actually being bound* by those norms.

At this point red flags should be springing up. Certainly we *cannot* felicitously perform *just any* speech act in the privacy of our own thoughts. This is because some speech acts require more to come off successfully than just the speaker's intention to perform them. Consider the case of *marrying two people*. A speaker cannot merely say silently to themselves "*I hereby pronounce you…*" and thereby make it the case that two individuals are married. The two individuals must be present and must have themselves formally and publically declared their intention to be married. The speaker herself must have the extrinsically determined status of a licensed officiant, and her pronunciation of

marriage must be audible and performed before witnesses. Or again, take the speech act of christening a ship. I cannot christen the Natural Environment Research Council's new polar research vessel merely by saying silently to myself "*I hereby christen you the 'RRS Boaty McBoatface'*." As the enthusiastic respondents of the NERC's poll found out—to their own chagrin—the consent of the owner of the vessel is ultimately also required for a successfully christening.[238] However, for the class of speech acts whose constitutive norms *may* be fully satisfied by a speaker with the right intentions, there is no obstacle to their also being performed in the privacy of the speaker's thoughts.

Another concern may be tugging at the corners of this discussion: every time we perform a speech act, must we always form an occurrent intention (of which we have full awareness) to mean what we say with the force of a particular speech act? And in order to do this, must we have fully and explicitly in mind exactly the constitutive norms of that speech act, taking ourselves to be bound by each and every one? Of course not! But this is just a general observation about what it is to perform a speech act according to Force Conventionalism, and not an issue unique to private or inner speech acts. *Meaning what one says* with a particular force does not always require a full-fledged awareness of the speech act type and its norms in inner speech any more than it does in public speech. Imagine uttering silently to yourself the sentence "*It will rain today*." We all know what it feels like to silently say such a thing and mean it as a *guess*. Imagine yourself thinking this as you gaze out at the ominous, gathering clouds with worried anticipation. We also all know how different it feels to utter that same sentence with the force of an *assertion*. Just imagine yourself looking out the window *at the rain currently falling* while you assert the

[238] BBC (2016).

same sentence silently to yourself in a tone of resignation (probably with a different word emphasis): "*It will rain today*."

In these examples, just as in public speech, it is not as though we must go through a mental warm up: "*I'd like to make an educated guess. Good. And such guesses require only a moderate level of justification—no outright belief or knowledge. Yes, that's the speech act for me. All right, get ready…here we go-----It will rain today."* I won't rule out the possibility that we occasionally proceed through such inner gymnastics. The worlds of private and inner speech can certainly be bizarre ones. We can easily surprise, shock, and confuse ourselves with the things we silently say. However, no theory *requiring* such intentional preambles could possess any plausibility. Fortunately, Force Conventionalism is no exception. The requirement that one intend one's utterance to possess a certain force, and thereby to bind oneself by the respective norms, is no more exotic than our everyday experience of *meaning what we say* in different ways.

To underline how effortless it is to engage in dialogic inner speech, consider the connection between inner speech and social speech. Because one's concepts of speech act types are constructed out of the social milieu in which they were first learned, the norms constitutive of each type are internalized often without ever having first been made explicit. The ability to speak, whether socially, privately, or internally, is more a matter of knowledge *how* than of knowledge *that*. This is why philosophical disagreement about the norms constituting particular speech act types cannot cast doubt on our pre-theoretical grasp of those norms and ability to bind ourselves by them—any more than a scientific disagreement about the best way to mathematically describe the physics involved in a child riding a bike can cast any doubt on the child's ability to ride the bike. As a matter of fact,

just as the actual movements of the child on the bike constrain the adequacy of scientific theories being built to describe them, our actual linguistic practices constrain the adequacy of our philosophical theorizing about them. The skill of speaking, just like the skill of bike riding, does not require explicit knowledge of every step one takes in the execution of that skill. Force Conventionalism recognizes this fact, even as it imposes a structured explanation that helps us understand how the process works.

I realize that the notion of being bound by a norm simply in virtue of taking oneself to be bound is an unusual one. This is why the analogy with solitary games is helpful. Just take single person card games. Suppose that, in the course of a game of Klondike, you accidentally turn over two cards instead of three during one of your passes through the deck. When you realize your mistake, as you will when your next pass through the deck suddenly reveals different cards, your feeling is one of *having broken a rule*—and this feeling is not misplaced. This is because part of *what it is* to play a solitaire card game like Klondike just is to *take oneself to be bound* by the rules of the game. And taking oneself to be bound by the rules of the game is sufficient for *actually being so bound.* If you "finish" the game despite your mistake, the experience of "winning" will possess at least some measure of hollowness.

Section 5: Appropriating Interpersonal Conventions for Intrapersonal Use

Though considering solitaire games is helpful, the analogy is ultimately imperfect. This is because such games are *originally designed for use by a single person*. In contrast, conventions like speech acts are originally designed for interpersonal use. This is an important point of disanalogy. If I am right that we can perform solitary analogues of social

speech acts, then such a use is a derivative one. Perhaps a more instructive analogy may be constructed using a game like chess.

Chess is a game originally designed for two players. Not only have variations arisen that allow for *more* than two players, but it is also certainly possible to play a game *against oneself*.[239] Imagine playing a game of solitary chess, switching sides every turn. If the idea seems odd to you at first, you may be in good company. Take the following scene from Netflix's show *Luke Cage*. When the police raid the barbershop where Cage usually hangs out and find only local chess master Bobby Fish inside, a half-finished chess game on the table presents itself as evidence that Cage may have recently been there too. Bobby Fish attempts to explain to the unimaginative police officers questioning him why a person might want to play a solitary game.

> DETECTIVE: "I understand playing against the computer but, uh, how do you beat yourself?"
> BOBBY: "Chess is a game of anticipation. If you know every single possible move you can anticipate your adversary's next move before he does."
> DETECTIVE: "Oh. [Sarcastically] That's deep. If we find out you've been hiding him, there'll be hell to pay!"[240]

Importantly, just like in the case of speech acts, there are reasons for playing chess that are not available to you when you are playing in a solitary context. As Haase points out with respect to solitary card games: "There is a sense in which I can win when I play Solitaire. But when I do, I have not beaten myself and have not been defeated by myself. There is no place for *winning against someone* in this game." But that does not mean that

---

[239] The most I have seen is retired logician Dale Holmes' decadent "Salmon P. Chess" variant that allows for ten players at a time. See Holmes (2013).

[240] *Marvel's Luke Cage* Season 1, Episode 12: "Soliloquy of Chaos."

you cannot *play* such games, it just means that you are limited in what you can reasonably try to accomplish *by playing them*. There are many reasons for playing chess that *do* make sense even when one is by oneself. Perhaps, like Bobby, you are a student of the game and are taking the opportunity to explore different strategies—how a certain defense holds up against a certain attack, etc. If you know every possible move that could be made against your chosen opening, "you can anticipate your adversary's next move before he does" in subsequent games against a real opponent.[241] Not only is it essential in such cases to understand oneself as subject to the rules in order to successfully perform the activity one is trying to perform, it is also essential in order to accomplish the goal one has set oneself. A series of illegal moves will do little to expand your understanding of the possible counters to your chess opening.

In chess, just as much as in the game of Klondike, taking yourself to be bound by the rules of the game is sufficient for *actually being bound* by them. In other words, all there is to playing chess is *intending to play chess* plus making the right sorts of movements with the right sorts of pieces.[242] Suppose that at a particularly critical point in the game you accidentally move one of the kings into check without noticing. When, several moves later, you realize the mistake you have two options: (1) return the board back to its pre-mistake state and continue with a legal move or (2) give up on your goal of playing a fair game of chess. It is certainly the case that you have broken a rule, just as much as if you were

---

[241] In perhaps a more pathological case, you might be playing chess against yourself in order to externalize and process through an internal struggle like the character of Elliot Alderson in USA Network's *Mr. Robot*. See Season 2, Episode 4: "eps2.2_init_1.asec."

[242] Even physical pieces are unnecessary if one has sufficient mental dexterity to imagine the game being played out within one's own head.

playing against a real opponent. This is because being constrained by the constitutive norms of chess is essential to your being able to conceive of yourself as playing that game in the first place. As long as you want to count as playing chess, you must follow the rules. Importantly, this obligation is merely conditional. There is nothing holding you to the rules besides your own desire to play chess. In the social case there are often *additional* norms in place. If you have agreed to play a chess game with me, I may criticize you for forfeiting the game. But even in the social case, *quitting* is importantly different from *cheating*.

Perhaps a critic will say, in a spirit sympathetic to the qualms about silent promises raised above, that it is part of their theory of chess that two or more players must be pitted against one another (even if one of them is only a computer rather than a human person). Perhaps, they say, there can be no chess without the possibility of *winning against someone*. Therefore, the critic concludes, whatever it is you are doing alone with that chess-board, it does not count as playing chess. My response is that the critic's theory is not a very useful one, and it is not true to our lived experiences with board games. A better analysis of chess would help us to distinguish the case of solitary chess from a case in which the chess pieces are being used to play a game of solitary checkers, and to distinguish both of these from a case in which a small child randomly slides the pieces around the board in imitation of her parents (even if her behavior happens to conform to the rules *by accident*). A theory that defines chess playing as acting in such a way as to intend to count as *assessable* by the rules of the game (whether alone or cooperatively with others) will be able to do just that.

In the case of speech acts, the analogous challenge would come from a theory that says that genuine performance of a speech act requires having a particular perlocutionary

intention.[243] Assertion requires intending to produce a belief in someone, asking a question requires intending that someone else give you information, making a promise requires intending someone else to hold you to it. But just as it is a confusion to identify illocutionary acts with their typical perlocutionary effects, so it is a confusion to require *intending to produce a particular effect* as a precondition for counting as having performed a speech act of a particular type. Such a theory would rule out the possibility of inner speech acts from the start, not to mention many instances of social speech acts performed for nonstandard reasons, and hence would not be very useful. A useful theory will help us to distinguish silent assertions from silent predictions, promises, questions, etc., and to distinguish all silent speech acts from silent acts of speech. This is exactly what a carefully refined version of Force Conventionalism allows us to do. In Part V we will see how further research at the intersection of Force Conventionalism, inner speech, and my own hybrid speech act theory could also serve to illuminate an important dimension of our normative thought.

To conclude this discussion, notice that there is another important respect in which the solitary chess game is analogous to performing an inner speech act. That is, in order to intend to play chess against oneself, it is not as though one must first rehearse every rule of chess and explicitly commit to following each of them. If you *know how* to play chess, all you have to do is *decide to play* chess—deciding to count as subject to the rules is implicit in this decision because they are constitutive of the activity upon which you take yourself to be embarking. In the same way, if you *know how* to perform a speech act like

[243] For some examples of theories in this vein see footnotes 221, 222, and 225 above.

assertion, all you have to do is *decide to mean* your utterance (silent or otherwise) with that force. Being subject to assessment according to the felicity conditions of the speech act is implicit in this decision because they are constitutive of the activity upon which you take yourself to be embarking. A speaker need not possess the terms, or even the precise concepts, that philosophers and linguists use to identify particular speech acts. In fact, it is quite implausible that most people would use the terms that experts do, but it is just as clear that they grasp the differences between different speech acts nonetheless. Just take this example from the *Star Trek* episode "That Which Survives." Upon discovering that the USS *Enterprise* has been suddenly transported nearly 1000 light years, first officer Spock and chief engineer Scott have the following conversation:

> SCOTT: But that's not possible. Nothing can do that.
> SPOCK: Mister Scott, since we are here, your statement is not only illogical but also unworthy of refutation. It is also illogical to assume that any explosion, even that of a small star going supernova, could have hurled us a distance of nine hundred and ninety point seven light years.
> SCOTT: The point is it shouldn't have hurled us anywhere. Why, it should have destroyed us immediately, vaporized us.
> SPOCK: That is correct, Mister Scott, by all the laws that we know. There was no period of unconsciousness. Our ship's chronometers registered a matter of only a few seconds. Therefore, we were displaced through space in some manner which I am unable to fathom.
> SCOTT: What you're saying is that the planet didn't blow up, and the captain and the others, they're still alive!
> SPOCK: Please, Mister Scott, restrain your leaps of illogic. I have said nothing. I was merely speculating.

When Spock says that he has "said nothing" it is clear that he does not mean that he has made no utterance. He also does not mean that he has merely performed an act of speech with no speaker meaning, falling short of speech act-hood. By claiming that he was "merely speculating," Spock clarifies the *force* with which he meant his utterance. He was

not performing the speech act Scott thought he was, but rather a less epistemically demanding one: making a conjecture. Even though Spock doesn't use the term 'assertion' when he says "I didn't *say* anything," it is clear that something like what philosophers would call "asserting" is what he is denying having done.

Here, then, is a summary of my argument. There is an important theoretical distinction between actually being bound by a norm and taking oneself to be bound by a norm. Acknowledging this distinction allows us to see that speakers who perform speech acts silently to themselves view themselves as subject to assessment by the speech acts' constitutive norms. Such a self-conception does not require an explicit theoretical grasp of speech acts and norms. Rather it involves a skill (knowledge how)—an ability to use a convention that is first learned by engaging with social speech. As social speech is gradually internalized to form silent speech, the constitutive norms of speech acts are internalized as well. Taking herself to be bound by the norms constitutive of a certain speech act type allows a speaker to *mean* what she says with the force of that speech act type.

According to force conventionalism, this is enough for a speaker to count as having actually performed a speech act, even if the utterance is an instance of private or inner speech. If the speaker successfully performs the speech act, its constitutive norms do apply to her. This manner of "binding oneself" sidesteps debates like that between Korsgaard, Darwall, and Haase. This is because the quality of the norms in question is not categorical, but only conditional on the desire to perform a speech act of the given type. Nothing requires you to play chess, but if you *want* to count as having played chess (rather than checkers or no game at all) then you *are* constrained by the rules defined by an originally

social convention. As it turns out, the distinction that was initially important for the purposes of understanding the argument ends up collapsing in the particular case at hand. When it comes to many speech acts, just as in solitary games of Klondike or chess, *taking oneself to be bound* can be sufficient for *binding oneself*. To be clear, there are speech acts such as marrying and christening that require considerably more cooperation from the world outside the speaker in order to come off successfully. However, for those speech acts requiring only that a speaker *mean* what they say with the right force (promising, asserting, guessing, questioning, etc.), one can bind oneself merely by intending to do so. Force Conventionalism is compatible with the existence of inner speech acts.

Part V: A Brief Survey of Normative Inner Speech

> I'm a good person. Let me tell you, I don't cheat on my wife. I don't cheat on my wife because one morning our whole family was in bed—the dogs, the cats, the girls, my wife, and we were just giggling and it was pure. It was perfect. And I thought to myself, "I don't ever want to screw this up. This is the most important thing, this is what life is about." And they got up to make chocolate chip pancakes. And I laid in bed and I said, "I will never cheat on my wife." I had a conversation with myself. I said, "If I ever get into a situation where a hot girl is flirting with me or I think she's flirting with me and it seems like it could go further, I'm just going to cock block myself. I'm just going to look her in the face in front of everyone and go, 'I don't cheat on my wife!'"
>
> Now, I may be wrong. She may not be hitting on me. She may throw a drink in my face, slap me, or I may be right. None of that matters to me. What matters to me is that I don't cheat on my wife because I've already had that conversation. Here's the problem: I never had that conversation about robbing trains.[244]

[244] *Bert Kreischer: The Machine* (2016).

We have already looked in depth at the case of public normative speech. When we make a statement with normative content and put it forward with the force of a particular normative speech act (moral, aesthetic, epistemic, rational, etc.), we express our belief in what we say as well as an appropriate motivational or affective attitude. Which particular attitudes are expressed depends on the normative terms we use, as well as the normative domain we are speaking within. We can utter sentences that have normative content and put them forward with non-normative force (ordinary assertions, guesses, suppositions for the sake of argument, etc.), or with no force at all (mere acts of speech). When we use normative language outside the normative domains of discourse, our speech may satisfy the goals we have set ourselves (e.g. to trace the logical relationships between concepts or to memorize lines for a play). It may be all we are capable of at the moment (because of a personality disorder, a recent tragedy, or an argument with a disturbing conclusion that appears sound). However, uses of normative language in non-normative speech acts are impoverished ones—they are disconnected from motive and affect and hence are severely limited in their ability to contribute to deliberation, coordination, and action.

What are we doing when we think about norms? It depends on what is meant by "thinking." If what we are trying to pick out are dispositions or subconscious processes (beliefs, desires, language of thought, etc.) then my theory has little to add. However, if we take a closer look at one mode of thinking that takes the form of a conscious activity, the hybrid speech act view has a great deal to say about what we are doing when we think about norms. In this last section I will trace out a rough sketch of how this works by drawing together several lines of previous arguments. My primary goal will be to demonstrate that the hybrid speech act view has the resources to respond to Mark

Schroeder's challenge raised at the end of the previous chapter, that hybrid views must abandon hope for consistency with broadly Humean theories of mind in order to maintain a form of motivational judgment internalism in the realm of normative thought.

Essentially Schroeder's challenge runs as follows. Commitment to motivational judgment internalism requires admitting that "people who genuinely believe moral claims will be motivated."[245] However, commitment to Humeanism requires maintaining that beliefs and desires are distinct mental states, and that no necessary connections hold between them. We have already seen that the hybrid speech act theory remains consistent with Psychological Humeanism (and the intuitions brought to bear by externalists) by allowing that it is possible for individuals to at the same time have a belief with normative content and fail to have any of the relevant motivations or affective responses.

In the realm of normative public speech, the hybrid speech act theory's commitment to motivational judgment internalism was cashed out in terms of the necessary connection between motive/affect and sincere normative speech act performance. If a speaker sincerely utters a normative statement with the force of a normative speech act, the speaker will both believe what she says and have the appropriate motivating or affective state. Is it possible to follow a similar strategy in the realm of normative thought?

Schroeder is right, but only because he has stacked the deck in his favor. It is true that Psychological Humeanism is straightforwardly incompatible with the claim that "people who believe moral claims will [always] be motivated." This is not a problem peculiar to hybrid views. It is merely the most recent statement of an observation made famous by Michael Smith in *The Moral Problem*: there is a contradiction between (1) the

[245] Schroeder (2014), p. 283.

claim that there are *no* necessary connections between beliefs and motivational states, and (2) the claim that there *is* a necessary connection between having a moral belief and being motivated.

The question of whether hybrid views are capable of making the contradictory suddenly consistent is not a very interesting one—hybrid views are not in the business of proposing alternative logics. A more interesting question is whether hybrid views give us the tools to develop a new version of motivational judgment internalism: one that still gives us everything we wanted from the old version, but this time in a way that is consistent with Humeanism. In the realm of public normative speech, the hybrid speech act theory does just that. What about the realm of normative thought?

In his book *Confusion of Tongues*, Stephen Finlay attempts to quiet suspicions about the difficulty of translating a theory of normative speech into a theory of normative thought:

> Despite the breadth of this conception of pragmatics, some will object it's still too narrow to be the key to solving metaethical puzzles, because whereas the domain of pragmatics is speech or utterances, the challenges cataloged above arise also in normative *thought* or judgments.…But these objections overlook important connections and parallels between speech and thought. On one hand, insofar as judgments are themselves linguistically formulated, they are plausibly internalizations of speech ("speaking to ourselves"), and therefore will resemble contributions to imagined conversations and inherit many pragmatically influenced features. It's implausible that although when speaking aloud we may omit words the audience can easily recover, when saying things to ourselves we must mentally token only fully explicit sentences: introspection finds similar patterns of ellipsis in thought as in speech.…While I'll seek to be sensitive to relevant differences between normative speech and thought, the following chapters will largely focus on speech to avoid having to say everything in two different ways.[246]

[246] Finlay (2014), pp. 120–121.

David Copp also sees the transition from theorizing about normative speech to theorizing about thought as relatively unproblematic:

> Rather, what this line of reasoning about the relation between moral discourse and moral thought suggests to me is that the thoughts of articulate beings are governed indirectly by norms that affect the meaning of the sentences and terms they would use to express these thoughts. In this sense the thoughts of articulate beings can have a linguistic aspect. A person who has an occurrent thought might entertain a sentence she could use to express it. This idea seems to be supported by the fact that when we vocalize a thought, we normally are able straightaway to assert it without needing to search for the words. It also seems to be supported by the fact, familiar to people who are bilingual, that there is a difference between, say, "thinking in French" and "thinking in English." If our thoughts can have a linguistic aspect in this way, then a person with an occurrent moral thought might entertain a sentence she could use to express this thought.[247]

These assurances, while suggestive, are hardly satisfying. The hybrid speech act theory can do better. It can do this by connecting the metanormative discussion up with the phenomenon of inner speech. Surely a great deal of our thought about norms takes place at a subconscious level and many of our mental states that take normative contents are dispositional states like beliefs or desires.

Even the conscious activity of thinking about norms might sometimes employ non-linguistic forms of thought like mental pictures, sounds, and other nonlinguistic occurrent mental episodes. However, for most people a substantial portion of their thinking about norms takes the form of inner speech. Earlier in this chapter I surveyed some of the recent empirical work on inner speech. There is clearly a great deal more research to be done, but already we may say with confidence that inner speech plays an important role in many

[247] Copp (2014), p. 58.

areas of human life from problem-solving and action planning, to self-awareness and social understanding.

Part of what enables inner speech to be so dynamic and effective is that it can take many different forms. Inner speech can come in both expanded and condensed varieties. It can involve mere acts of speech, as when we silently repeat a phone number or items on a grocery list to remember them. It can involve imagining speech acts being performed, as when you silently play back a past conversation or rehearse one that could possibly occur. It can also involve us performing speech acts, by meaning what we silently say with the force associated with the appropriate public speech act. As argued in Part III, the speech act theory of force conventionalism has the resources to explain how it is that we are able to perform speech acts when no audience is present: by speaking with the intention of being subject to the felicity conditions constitutive of those speech acts.

The hybrid speech act view adds to this story its list of unique and previously unrecognized types of speech acts, the various normative speech acts. When we think about norms, often what we are doing is speaking silently about normative subjects. When we do so from inside one of the particular normative domains we mean what we silently say with the force of the particular normative speech act in question (moral, epistemic, aesthetic, prudential, etiquette, etc.). We are able to do this by intending to bind ourselves by the felicity conditions constitutive of that particular type of normative speech act. Because we are speaking under the appropriate force, we express (silently) that we have satisfied the appropriate felicity conditions of belief and motive or affect. If, for example, someone *sincerely* makes a moral claim in inner speech with the force of a moral speech act, then

they both believe what they (silently) say and are at least somewhat motivated to act accordingly.

The form of judgment internalism that the hybrid speech act theory endorses in the realm of normative thought is parallel to that endorsed in normative public speech: if a thinker sincerely (and silently) utters a normative statement in inner speech with the force of a normative speech act, the thinker will both believe what she says and have the appropriate moving state. This is perfectly consistent with Psychological Humeanism. A thinker may *believe* some normative claim without possessing the appropriate motive/affect. A thinker may even utter a normative claim in inner speech and have no relevant motivation or affective response whatsoever, but to that extent her normative speech act would be insincere. Such a case would be like those discussed above when you silently assert to yourself something that you come to discover that you do not believe. As in public speech, so in inner speech—we do not always satisfy the felicity conditions of the speech acts we perform.

Though it is consistent with Humeanism, this form of motivational judgment internalism still preserves everything we wanted from the original. A thinker whose inner speech about normative topics is performed from outside any of the normative domains of silent discourse is thinking in an impoverished way. Their thought will be severely limited in its ability to help them deliberate, coordinate, and act. In contrast, thinking about normative topics from within a normative domain demands our practical engagement.

Often what happens in cases of deliberation is that our normative judgments *sotto voce* (e.g. 'eating dessert would be foolish!') can ring hollow—uncovering our own cognitive, motivational, or affective disunity. This can prompt us to act on the basis of

countervailing impulses ('but it looks so delicious') or to work towards a more harmonious inner state. Sometimes this means revising our beliefs ('perhaps eating it wouldn't be *so* bad'), sometimes our intentions ('I'll just exercise more later'), and sometimes working to suppress our desires ('I must stand strong. Remember how guilty I felt last time?'). Performing a normative claim in the privacy of one's own thoughts makes the sincerity conditions of that speech act salient. This pushes deliberation forward by uncovering the various ways in which one might not be able to wholeheartedly endorse the claim, and the variety of reasons that could be brought to bear in favor or against it.

The excerpt quoted above from Bert Kreischer's stand-up comedy routine *The Machine* illustrates what can happen when we make fully sincere, wholehearted normative speech acts in the privacy of our own thoughts.[248] The character in Kreischer's story recalls a moment when he silently states to himself that his family is a source of great value ("This is the most important thing, this is what life is about"). For this claim to be sincere, Bert must both believe what he says and also possess the appropriate moving attitudes. The fact that he does have these attitudes is vividly raised to his attention and voiced in another silent statement—"I don't ever want to screw this up." In turn this pushes his deliberation forward. He forms a strategy for handling situations in which he might be tempted to act in ways that could undermine the value he has identified. This event of deliberation has had practical effects in his life ("I don't cheat on my wife *because* I've already had that conversation"), and contributes to his conception of himself as "a good person."

[248] It is unclear from the way Kreischer tells the story whether these private statements are uttered aloud or silently. What is important is that we can easily imagine all of this taking place silently, so that is the way I will phrase my discussion. After all, Kreischer's Russian Mafia story is likely (one hopes) at least partly fictional.

In contrast, the fact that Bert has had no such conversation with himself about robbing trains means that he is at a disadvantage when he is faced with that situation. Of course, this is a stand-up comedy routine and the humor arises in part from the fact that most of us would not find ourselves at a disadvantage were *we* to be faced with such a situation. The proposition of robbing a train just doesn't feel like the sort of dilemma, the adequate handling of which one needs a memory of rich deliberation to fall back on.

We don't need to think hard to prepare ourselves for the easy cases. However, we can easily imagine being faced with moral challenges for which we are unprepared. Part of what it is to *be unprepared* is to lack a memory of having settled the question. Take the example of cheating on one's taxes. Suppose you haven't really thought much about the ethics of paying taxes (you try not to think about taxes much at all!), so when you are suddenly faced with a dilemma you are at a disadvantage: "If I declare this particular bit of income, I will have to pay back *a lot* of money. But if I don't, I'll receive a refund."

Your disadvantage has at least two dimensions. The first is that at no point in the past have you taken the time, *at a distance from any particular tempting situation*, to silently give voice to the thought "cheating on one's taxes is wrong." If you had said such a thing, and had your utterance rang sincere, it would have raised to the level of conscious awareness both your belief that cheating is wrong and your motivation not to do such a thing. The fact that you have not gone through this deliberative process means that you do not have the memory of being occurrently aware of these clear-eyed dispositions in yourself to draw on in the moment of crisis.

Of course, you may still be able to deliberate in the moment of temptation, but you are at a disadvantage because *precisely what* makes the situation a tempting one is the fact

that it pits competing desires, values, and interests against one another. A memory of having settled the question previously, with its accompanying impression of psychological unity, would have better positioned you to cut through the motivational morass in which you find yourself.

The second, related respect in which you are at a disadvantage is that you have no strategy to fall back upon to handle the temptation you feel. If you *had* deliberated about the immorality of cheating on your taxes, and suspected that you might find yourself in such a situation, then your awareness of your desire *not to give in* to temptation might have prompted you to establish a strategy to get yourself out of it (just like Bert planned for instances of temptation to marital infidelity). Without such a strategy to fall back on, in the moment of temptation you may find yourself acting in a way that you come to disown later upon further reflection.

To sum up: Normative inner speech, when performed within normative domains of discourse, is uttered with the force of normative speech acts. These silent speech acts inherit their constitutive norms from their public counterparts. When we perform normative inner speech acts, we indicate to ourselves that we satisfy the sincerity conditions of those speech acts—we express the relevant cognitive, affective, and/or motivational states. This makes the sincerity conditions salient and, often, causes us to become aware that we do or do not in fact possess those appropriate mental states. Self-awareness of psychological harmony with respect to a normative claim can prompt us to form strategies and resolutions for following through on our values in difficult situations. Self-awareness of psychological disunity with respect to a normative claim can prompt us to deliberate about revising our beliefs, sensitizing our affective responses, and/or modifying our motives and behavior.

CHAPTER 5

CONCLUSION AND SUGGESTIONS FOR FURTHER RESEARCH

Summary of Project

What are we doing when we think and talk about norms? The form of the question itself suggests the beginning of an answer: we are *doing* something. That is to say, when we think and talk about norms we are first and foremost engaged in activities. One thing that I hope has emerged from the discussions of this dissertation is that the most felicitous instances of these activities involving *using* norms to do things. When we speak and think about norms *from within* normative domains of discourse, what we are doing is not merely engaging in activities that are *about* norms, but engaging in projects that put norms to work—to deliberate about our values, express our inner states, give life to our motivations, stir up our sentiments, and bring ourselves to act—both collectively and individually.

In the first chapter, I described the hybrid metanormative program as one that strives to take seriously both the practical and the cognitive dimensions of our normative activities. It aims to explain what we are up to, without taking the easy way out of simply positing hybrid mental states or sui generis necessary connections. In the past theories in

this tradition have usually proceeded by emphasizing one dimension as fundamental and seeking to appease our intuitions with respect to the other. I argued that such a strategy is not attractive in its own right, but only as a form of compromise if no fully hybrid alternative is possible. Before theorizing takes hold, the most effortless way of interpreting our full-throated normative speech and thought is as putting equally at stake the demands of both cognitive and practical dimensions. On the one hand, it is natural to think that part of what it *is* for thought and talk to count as normative *just is* for its serious application to be expressive of motivational or affective states. On the other hand, we also rely heavily on the presumption that individuals who make normative claims believe what they are saying—just as speakers who make non-normative assertions do. If it is possible to hang on to both aspects of normativity, then we would like to do so.

In the second chapter I examined two prominent theories that come very close to capturing both aspects of our normative thought and speech. Though they ultimately fail, their manner of failure is instructive. Part of what it is to deal justly with the practicality of our normative discourse is to recognize that speakers cannot cancel implicatures of motive or affect without clarifying that their use was not intended to be a normative one in the first place. It is in this respect that Stephen Finlay's End-Relational Theory falls short. Along the way he also commits himself to a particularly controversial theory of the semantics of normative terms. At the same time, it is important that normative language and thought do not involve communications of motive and affect that are capable of scoping out of embedded contexts. Despite its practicality, normativity lacks the charged element that is present in certain kinds of colored language like slurs and epithets (and possibly conventional implicatures). It is not normative language *simpliciter* that is practical, but

peculiarly normative *uses* of language. On this score David Copp's Realist Expressivism misses the mark. What emerges is a challenge to future attempts to construct hybrid theories: to construct an explanation of normative language according to which only certain uses are practical, but from whose practicality it is also difficult to untangle oneself.

The third chapter sought to answer this challenge in the realm of public speech by gathering insights from speech act theory and using them to clarify what it *would be* to posit a genuinely normative use of language. Especially important is the idea that speech act types are individuated by their constitutive rules or "felicity conditions." Once this is recognized, it is clear that if there *are* speech acts that requires both cognitive and practical elements of felicitous speakers, they must be of previously unrecognized types. The achievement of this approach to the normative use of language consists in its putting both elements on exactly equal footing for the first time. The mechanism by which a normative speaker expresses her belief is precisely the same as the mechanism by which she expresses her motivational or affective state. Because the form of activity in which she is engaged is one that requires both attitudes in order to count as sincere, any deficit along either dimension is sufficient to disqualify her from wholeheartedness. While it is possible to speak about normative topics with the force of speech act types that have less demanding requirements, such conversations severely limit what speakers are able to *do with* norms. Fully normative uses of language, precisely because of the demands they make, enable speakers to do much more—to explore, question, debate, express, coordinate, and act upon the conjunction of their cognitive and practical commitments.

While the construction of a fully hybrid theory of the public normative use of language is a considerable accomplishment, the ultimate prize consists in being able to cash

in the explanation of public speech for a glimpse into normative thought. In chapter 4 I began by making explicit the Vygotskian assumptions to which such a project is committed. In order for a theory of normative public speech *to be translatable* into a theory of normative thought, there must be a mode of thought that (1) takes place in natural language, (2) inherits its structure from public speech, and (3) plays a significant cognitive role (is not merely epiphenomenal). After a survey of some of the empirical work into the phenomenon of inner speech, it emerged that we have good reason to think that there is a mode of thought that does satisfy (1)–(3), but only if we can make sense of the idea of its sometimes taking place in a distinctively *dialogic* form. This led to an argument for the coherence of truly solitary performances of speech acts. Drawing the various strands of argument together, I concluded that parallel to—and dependent on—the normative use of language there is also a *normative use of thought*.

When a thinker engages in dialogic inner speech from within one of the distinctively normative domains, her activity is only intelligible from her first person point of view as being of the type she intends it to be if she understands herself to count as evaluable by the rules constitutive of her chosen activity. Just as I can only play a solitary game of chess if I take deviation from the rules to count as *rule-breaking*, I can only engage in the normative use of thought if I take deficits along either cognitive or practical dimensions to count as failures of full sincerity. As in public speech, so also in thought: the normative use allows us to *do things* with norms—to explore, debate, question, express, coordinate, revise, plan, and act upon our values. Self-awareness of psychological harmony with respect to a normative claim can prompt us to form strategies and resolutions for following through on our values in difficult situations. Self-awareness of psychological

disunity with respect to a normative claim can prompt us to deliberate about revising our beliefs, sensitizing our affective responses, and/or modifying our motives and behavior.

Suggestions for Further Research

In many ways the discussion of the hybrid speech act theory offered in this dissertation represents the beginning rather than the end of a program of inquiry. The explanation given of the way normative language operates in thought and public speech raises many interesting questions, and has the potential to provide insight into a diverse array of philosophical and scientific issues. To mark this transition from *what has been done* to *what there is yet to do*, I will discuss just a few of these opportunities for further research.

The first is in the area of philosophy of language. In the course of describing and defending the hybrid speech act theory I considered the questions of how to individuate speech act types, how to decide when to posit a new speech act type, and how to define what it takes to count as having performed a speech act. A better understanding of the answers to these three questions can help us to analyze the differences between many different uses of language, some of which have been discussed thoroughly and others which have seen virtually no discussion at all. For example, debates in epistemology about the speech act of assertion would be greatly served by establishing a common ground on the nature and individuation conditions of speech acts.[249] Other forms of speech—some

[249] See for example Goldberg (2015) and McKinnon (2015).

that have seen less philosophical attention than *assertion* has—seem to involve a requirement that the speaker be less than fully sincere. Take for example speech that aims at irony, sarcasm, bullshit, and gaslighting. Investigating whether some of these uses of language involve hitherto unrecognized species of speech acts could help us to understand their expressive and sometimes even coercive power.

A better understanding of what it takes to perform a speech act, and particularly of what it is to speaker-mean without an audience present can also help us to get a grip on uses of language in social media. Facebook posts, Tweets, online blogs and the like usually involve speech acts and not mere acts of speech. Most of the time a speaker knows she will be heard, but not exactly who her audience will be. Sometimes a speaker may be unsure whether a particular social media utterance will find any audience at all. How do these facts interfere with or modify the norms and dynamics at play in online "conversations"? How does the possibility of commenting anonymously distort our ability to stick our necks out and express our inner states? Is trolling a previously unrecognized type of speech act? If so, what are its constitutive norms? Precise methodology in speech act theory has the potential to contribute to our understanding of the function, dysfunction, expressive power, and manipulative potential of a host different forms of communication.

The second opportunity has to do with the scientific exploration of inner speech. Only recently have psychologists working on inner speech started carefully distinguishing between dialogic and non-dialogic forms of inner speech. This distinction could be made even more precise with concepts from speech act theory, especially by investigating the difference between *imagining* a silent speech act and *performing* a silent speech act. How is it that the execution of different types of speech acts involves the performance of

different cognitive functions? Asserting and questioning in inner speech obviously have different roles, but what about the difference between making a conjecture and supposing for the sake of argument? What about the difference between normative speech acts and non-normative speech acts? By making explicit our best understanding of the constitutive norms of various speech act types, philosophical theory and empirical research can test and guide one another. Experiments can investigate precise felicity conditions to refine our understanding of speech act norms. Philosophy can uncover implicit rule following that explains the functional roles different inner speech acts are able to play, and the ways in which different types of speech acts can interact with one another in inner conversations. Drawing out the connections between public speech acts and inner speech acts opens up a new avenue for language acquisition research to contribute to research in cognitive development.

There are also important implications for responsible methodology in normative philosophy. When inquiring into normative questions in ethics, epistemology, aesthetics, etc., philosophers should be clear about the scope and purpose of their investigations. If their aim is a truly normative one, then their discussion must be oriented towards more than just knowledge in the traditional epistemic sense—it should result in patterns of feeling, motivation, and action as well. This means that assessing claims made within a normative discussion requires evaluating more than just their truth, justification or logical relationships with other statements. It also demands assessing the affective and motivational states of those putting the claims forward, and our own as we consider them. The hybrid speech act theory casts normative commitment as a state of being that involves different aspects of an agent's psychology aligning in harmony. Such a picture insists that

we reassess what it is to be sincere, when it is that we still have more deliberating to do, and the prejudices that may lead us to privilege some psychological states (like beliefs) over others (like emotions) in coming to practical conclusions.

Finally, the work in this dissertation opens up opportunities at the intersection of philosophy and psychology. For the same reason that the hybrid speech act theory has implications for the activity of normative philosophy, it can also help us to better understand ethical development, education, deliberation, and action. The hybrid speech act view emphasizes the status of social discourse as fundamental in the development of our ability to think through our normative commitments. Public conversations serve as the source of our normative concepts and of our understanding of the rules of normative speech. This both supports and helps to illuminate the work of scholars who see moral functioning as socially mediated, and the moral self as dialogic.[250] Conscious, intentional, normative cogitation is an internalized form of public normative dialogue. This means that theories of social speech can help us understand the way we think. It also means that discovering solutions to especially difficult problems may require conversations with others.

[250] Tappan (1990, 1997, 2006a, 2006b).

Bibliography

Alston, William 2000. *Illocutionary Acts and Sentence Meaning*. Ithaca, NY: Cornell University Press.

Anscombe, G. E. M. (1957). Intention. Oxford: Basil Blackwell.

Arendt, Hannah (1961). *Between Past and Future: Six Exercises in Political Thought*. New York: The Viking Press.

Austin, J. L. (1962). *How to Do Things with Words*. London: Oxford University Press.

Avramides, Anita (1989). *Meaning and Mind: An Examination of the Gricean Account of Language*. Cambridge, MA: The MIT Press.

Aydede, Murat (2015) "The Language of Thought Hypothesis." *The Stanford Encyclopedia of Philosophy* (Fall 2015 Edition), Edward N. Zalta (ed.). URL = <https://plato.
stanford.edu/archives/fall2015/entries/language-thought/>.

Ayer, A. J. (1936). *Language, Truth, and Logic*. London: Penguin Books.

Baddeley, Alan D. and Graham Hitch (1974). "Working Memory." *The Psychology of Learning and Motivation* Vol. 8, pp. 47–89.

Bach, Kent (1999). "The Myth of Conventional Implicature." *Linguistics and Philosophy* Vol. 22, pp. 327–366.

Bach, Kent & Robert M. Harnish (1979). *Linguistic Communication and Speech Acts*. Cambridge, MA: The MIT Press.

Barker, Stephen (2000). "Is Value Content a Component of Conventional Implicature?" *Analysis* Vol. 60, pp. 268–279.

Barker, Stephen (2014). "Pure versus Hybrid Expressivism and the Enigma of Conventional Implicature." In *Having It Both Ways: Hybrid Theories and Modern Metaethics*, G. Fletcher and M. Ridge (eds.). New York: Oxford University Press, pp. 199–222.

Bar-On, Dorit (2004). *Speaking My Mind: Expression and Self-Knowledge*. New York: Oxford University Press.

Bar-On, Dorit and Matthew Chrisman (2009). "Ethical Neo-Expressivism." In *Oxford Studies in Metaethics Vol. 4*, R. Shafer-Landau (ed.). Oxford: Oxford University Press, pp. 133–166.

Bar-On, Dorit, Matthew Chrisman, and James Sias (2014). "(How) Is Ethical Neo-Expressivism a Hybrid View?" In *Having It Both Ways: Hybrid Theories and Modern Metaethics*, G. Fletcher and M. Ridge (eds.). New York: Oxford University Press, 223–247.

BBC (2016). "'Boaty McBoatface' polar ship named after Attenborough." *BBC News*, May 6, 2016. URL = <http://www.bbc.com/news/uk-36225652>.

Bell, David (1979). *Frege's Theory of Judgment*. Oxford: Clarendon Press.

Bellah, Robert N., Richard Madsen, William M. Sullivan, Ann Swidler, and Steven M. Tipton (1985). *Habits of the heart: Individualism and commitment in American life*. Berkeley, CA: University of California Press.

Bennett, Jonathan (1976). *Linguistic Behaviour*. New York: Cambridge University Press.

Bermúdez, José. L. (2003). *Thinking without Words.* Oxford: Oxford University Press.

Bert Kreischer: The Machine. Directed by Ryan Polito. Written by Bert Kreischer. Showtime, 2016.

Björklund, Fredrik, Gunnar Björnsson, John Eriksson, Ragnar Francén Olinder, & Caj Strandberg (2012). "Recent Work on Motivational Internalism." *Analysis* Vol. 72, pp. 124–137.

Blachowicz, James (1997). "The Dialogue of the Soul with Itself." *Journal of Consciousness Studies* Vol. 4, pp. 485–508.

Blackburn, Simon (1984). *Spreading the Word*. New York: Oxford University Press.

Blackburn, Simon (1988). "Attitudes and Contents." *Ethics* Vol. 98, pp. 501–517.

Blackburn, Simon (1993). "Realism, Quasi or Queasy?" In *Reality, Representation, and Projection*, J. Haldane and C. Wright (eds.). Oxford: Oxford University Press, pp. 365–384.

Blackburn, Simon (1998). *Ruling Passions*. New York: Oxford University Press.

Blome-Tillmann, Michael (2008). "Conversational implicature and the cancellability test." *Analysis* Vol. 68, pp. 156–160.

Boisvert, Daniel R. (2008). "Expressive Assertivism." *Pacific Philosophical Quarterly* Vol. 89, pp. 169–203.

Boisvert, Daniel R. (2014). "Expressivism, Nondeclaratives, and Success-Conditional Semantics." In *Having It Both Ways: Hybrid Theories and Modern Metaethics*, G. Fletcher and M. Ridge (eds.). New York: Oxford University Press, pp. 22–50.

Borg, Emma (2004). *Minimal Semantics*. New York: Oxford University Press.

Borg, Emma (2012). *Pursuing Meaning*. New York: Oxford University Press.

Borge, Steffen (2009). "Conversational Implicatures and Cancellability." *Acta Analytica* Vol. 24, pp. 149–154.

Borge, Steffen (2013). "Talking to Infants: A Gricean Perspective." *American Philosophical Quarterly* Vol. 50, pp. 423–428.

Brandom, Robert (1983). "Asserting." *Noûs* Vol. 17, pp. 637–650.

Brandom, Robert (2009). *Reason in Philosophy: Animating Ideas*. Cambridge, MA: Belknap Press.

Brown, Jessica & Herman Cappelen (2011). *Assertion: New Philosophical Essays*. New York: Oxford University Press.

Burnyeat, Myles (1990). *The Theatetus of Plato*. Indianapolis, IN: Hackett Publishing.

Cappelen, Herman and Ernest Lepore (2005). *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*. Malden, MA: Blackwell Publishing Ltd.

Carruthers, Peter (1996). *Language, Thought, and Consciousness: An Essay in Philosophical Psychology*. Cambridge: Cambridge University Press.

Carruthers, Peter (2002). "The Cognitive Functions of Language." *Behavioral and Brain Sciences* Vol. 25, pp. 657–726.

Carruthers, Peter (2006). *The Architecture of Mind*. Oxford: Oxford University Press.

Carruthers, Peter and Bénédicte Veillet (2011). "The Case Against Cognitive Phenomenology." In *Cognitive Phenomenology*, T. Bayne and M. Montague (eds.). Oxford University Press, pp. 35–56.

Cholbi, Michael (2009). "Moore's Paradox and Moral Motivation." *Ethical Theory and Moral Practice* Vol. 12, pp. 495–510.

Cholbi, Michael (2011). "Depression, Listlessness, and Moral Motivation." *Ratio* Vol. 24, pp. 28–45.

Clowes, Robert (2007). "A self-regulation model of inner speech and its role in the organization of human conscious experience." *Journal of Consciousness Studies* Vol. 14, pp. 59–71.

Cohon, Rachel (2008). *Hume's Morality: Feeling and Fabrication*. New York: Oxford University Press.

Copp, David (1995). "Moral Obligation and Moral Motivation." In *On the Relevance of Metaethics*, J. Couture and K. Nielsen (eds.). University of Calgary Press, pp. 187–220.

Copp, David (2001). "Realist-Expressivism: A Neglected Option for Moral Realism." In *Moral Knowledge*, E. Paul, F. Miller, Jr., and J. Paul (eds.). New York: Cambridge University Press, pp. 187–220.

Copp, David (2009). "Realist-Expressivism and Conventional Implicature." In *Oxford Studies in Metaethics Vol. 4*, R. Shafer-Landau (ed.). Oxford: Oxford University Press, pp. 167–202.

Copp, David (2014). "Can a Hybrid Theory Have It Both Ways? Moral Thought, Open Questions, and Moral Motivation." In *Having It Both Ways: Hybrid Theories and Modern Metaethics*, G. Fletcher and M. Ridge (eds.). New York: Oxford University Press, pp. 51–74.

Cuneo, Terence (2014). *Speech and Morality: On the Metaethical Implications of Speaking*. New York: Oxford University Press.

Darwall, Stephen (2006). *The Second-Person Standpoint: Morality, Respect and Accountability*. Cambridge, MA: Harvard University Press.

Dien, Dora S.-F. (1982). "A Chinese perspective on Kohlberg's theory of moral development." *Developmental Review* Vol. 2, pp. 331–141.

Dummett, Michael (1973). *Frege: Philosophy of Language*. New York: Harper and Row Publishers.

Dreier, James (1999). “Transforming Expressivism.” *Noûs* Vol. 33, pp. 558–572.

Ehrich, John F. (2006). “Vygotskian Inner Speech and the Reading Process.” *Australian Journal of Educational & Developmental Psychology* Vol. 6, pp. 12–25.

Eriksson, John (2015). “Explaining Disagreement: A Problem for (Some) Hybrid Expressivists.” *Pacific Philosophical Quarterly* Vol. 96, pp. 39–53.

Fargo. Directed by Joel Coen and Ethan Coen. London: Polygram Film Entertainment, 1996.

Fargo. Created by Noah Hawley. Based on the film by Joel Coen and Ethan Coen. First released on FX, April 15, 2014.

Fernyhough, Charles (2004). “Alien voices and inner dialogue: towards a developmental account of auditory verbal hallucinations.” *New Ideas in Psychology* Vol. 22, pp. 49–68.

Fernyhough, Charles (2008). “Getting Vygotskian about theory of mind: Mediation, dialogue, and the development of social understanding.” *Developmental Review* Vol. 28, pp. 225–262.

Fernyhough, Charles (2009). “Dialogic Thinking.” In *Private Speech, Executive Functioning, and the Development of Verbal Self-Regulation*, Winsler, Fernyhough, and Montero (eds.). New York: Cambridge University Press, pp. 42–52.

Fernyhough, Charles (2010). “Vygotsky, Luria, and the Social Brain.” In *Self- and Social-Regulation: Exploring the Relations between Social Interaction, Social Cognition, and the Development of Executive Functions*, B. Sokol, U. Müller, J. Carpendale, A. Young, and G. Iarocci (eds.). New York: Oxford University Press, pp. 56–79.

Fernyhough, Charles (2016). *The Voices Within: The History & Science of How We Talk to Ourselves*. New York: Basic Books.

Finlay, Stephen (2004). ‘‘The Conversational Practicality of Value Judgment.’’ *The Journal of Ethics* Vol. 8, pp. 205–223.

Finlay, Stephen (2005). “Value and Implicature.” *Philosophers Imprint* Vol. 5: <http://www.
philosophersimprint.org/005004/>.

Finlay, Stephen (2014). *Confusion of Tongues: A Theory of Normative Language*. New York: Oxford University Press.

Fletcher, Guy and Michael Ridge (2014). "Introduction." In *Having It Both Ways: Hybrid Theories and Modern Metaethics*, G. Fletcher and M. Ridge (eds.). New York: Oxford University Press, pp. ix–xvi.

Fricker, Miranda (2009). *Epistemic Injustice: Power & the Ethics of Knowing*. New York: Oxford University Press.

Gaita, Raimond (2004). *Good and Evil: An Absolute Conception*. New York: Routledge.

Gale, Richard M. (1971). "The Fictive Use of Language." *Philosophy* Vol. 46, pp. 324–40.

Game of Thrones. "A Golden Crown." Season 1, Episode 6. Directed by Daniel Minahan. Story by D. Benioff and D. B. Weiss. Teleplay by J. Espenson, D. Benioff, and D. B. Weiss. Home Box Office, May 22, 2011.

Geach, Peter (1960). "Ascriptivism." *The Philosophical Review* Vol. 69, pp. 221–225.

Geach, Peter (1965). "Assertion." *The Philosophical Review* Vol. 74, pp. 449–465.

Geuss, Raymond (1996). "Morality and Identity." In *The Sources of Normativity*, Onora O'Neill (ed.). New York: Cambridge University Press.

Gibbard, Alan (1990). *Wise Choices, Apt Feelings*. Cambridge, MA: Harvard University Press.

Gilligan, Carol (1982). *In a different voice: Psychological theory and women's development*. Cambridge, MA: Harvard University Press.

Gilligan, Carol (1983). "Do the social sciences have an adequate theory of moral development?" In *Social Science as Moral Inquiry*, N. Haan, R. Bellah, P. Rabinow, and W. Sullivan (eds.). New York: Columbia University Press.

Gilligan, Carol and Jane Attanucci (1988) "Two moral orientations: Gender differences and similarities." *Merrill-Palmer Quarterly* Vol. 34, pp. 223–237.

Goffman, Erving (1978). "Response Cries." *Language* Vol. 54, pp. 787–815.

Green, Mitchell (2007). *Self-Expression*. New York: Oxford University Press.

Green, Mitchell (2015). "Speech Acts." *The Stanford Encyclopedia of Philosophy* (Summer 2015 Edition), Edward N. Zalta (ed.). URL = <http://plato.stanford.edu/archives/sum2015/entries/speech-acts/>.

Green, Mitchell (Forthcoming). "A Refinement and Defense of the Force/Content Distinction." In *New Work on Speech Acts*, D. Harris, D. Fogal, and M. Moss (eds.). Oxford: Oxford University Press.

Green, Mitchell & John Williams (2007). "Introduction." In *Moore's Paradox: New Essays on Belief, Rationality, and the First Person*, M. Green and J Williams (eds.). Oxford: Oxford University Press, pp. 3–36.

Grice, Paul (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Haase, Matthias (2014). "Am I You?" *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action*, DOI: 10.1080/13869795.2014.949065, pp. 1–14.

Hay, Ryan J. (2013). "Hybrid Expressivism and the Analogy between Pejoratives and Moral Language." *European Journal of Philosophy* Vol. 21, pp. 450–74.

Holmes, Dale (2013). "Salmon P. Chess." *The Chess Variant Pages*, April 25, 2013. URL = <http://www.chessvariants.com/contests/10/spc.html>.

Hopkins, Robert (2011). "How to be a Pessimist about Aesthetic Testimony." *Journal of Philosophy*, Vol. 108, pp. 138–157.

Huebner, Ann M. and Andrew C. Garrod (1993). "Moral reasoning among Tibetan monks: A study of Buddhist adolescents and young adults in Nepal." *Journal of Cross-Cultural Psychology* Vol. 24, pp. 167–185.

Jary, Mark (2010). *Assertion*. London: Palgrave Macmillan.

Jones, Simon and Charles Fernyhough (2007). "Neural Correlates of Inner Speech and Auditory Hallucinations: A Critical Review and Theoretical Investigation." *Clinical Psychological Review* Vol. 27, pp. 140–154.

Jowett, Benjamin, trans. (1952) "Meno." In *Great Books of the Western World Volume 6: Plato*, M. J. Adler (ed.). Chicago: Encyclopædia Britannica Inc., pp. 174–190

Joyce, Richard (2002). "Expressivism and Motivational Internalism." *Analysis* Vol. 62, pp. 336–344.

Kissine, Mikhail (2013). *From Utterances to Speech Acts*. New York: Cambridge University Press.

Korsgaard, Christine (1986). "Skepticism about Practical Reason." *The Journal of Philosophy* Vol. 83, pp. 5–25.

Korsgaard, Christine (2007). "Autonomy and the Second Person Within: A Commentary on Stephen Darwall's *The Second-Person Standpoint*." *Ethics* Vol. 118, pp. 8–23.

Korsgaard, Christine (2009). *Self-constitution: Agency, Identity and Integrity*. Cambridge, MA: Harvard University Press.

Kukla, Rebecca (2014). "Performative Force, Convention, and Discursive Injustice." *Hypatia* Vol. 29, pp. 440–457.

Mackie, J. L. (1977). *Ethics: Inventing Right and Wrong*. New York: Penguin Books.

Martin, Wayne M. (2006). *Theories of Judgment: Psychology, Logic, Phenomenology*. New York: Cambridge University Press.

Martínez-Manrique, Fernando and Agustin Vicente (2010). "What the…! The Role of Inner Speech in Conscious Thought." *Journal of Consciousness* Studies Vol. 17, pp. 141–67.

Martinich, A. P. (1984). *Communication and Reference*. New York: Walter de Gruyter & Co.

Marvel's Luke Cage. "Soliloquy of Chaos." Season 1, Episode 12. Directed by Phil Abraham. Written by Akela Cooper and Charles Murray. Netflix, September 30, 2016.

McDowell, John, trans. (1973). *Plato: Theaetetus*. New York: Oxford University Press.

Miller, Christian (2008). "Motivational Internalism." *Philosophical Studies* Vol. 139, pp. 233–255.

Millikan, Ruth (1998). "Language Conventions Made Simple." *Journal of Philosophy* Vol. 95, pp. 161–80.

Moon, Seung-Jae and Björn Lindblom (2003). "Two experiments on oxygen consumption during speech production: vocal effort and speaking tempo." In *Proceedings of the XVth International Congress of Phonetic Sciences* (Barcelona).

Morgan, Andrew (2017), "Solving the Puzzle of Aesthetic Assertion." *Southwest Philosophy Review* Vol. 33.

Morgan, Andrew (Forthcoming). "Hybrid Speech Acts: A Theory of Normative Thought and Language That 'Has It Both Ways.'" *European Journal of Philosophy* Early View, pp. 1–23.

Morin, Alain (2005). "Possible Links between Self-Awareness and Inner Speech: Theoretical Background, Underlying Mechanisms, and Empirical Evidence." *Journal of Consciousness Studies* Vol 12, pp. 115–134.

Morin, Alain (2006). "Levels of consciousness and self-awareness: A comparison and integration of various neurocognitive views." *Consciousness and Cognition* Vol. 15, pp. 358–371.

Morin, Alain (2008). "Self-awareness deficits following loss of inner speech: Dr. Jill Bolte Taylor's case study." *Consciousness and Cognition* Vol. 18, pp. 524–529.

Morin, Alain (2012). "Inner Speech." In *Encyclopedia of Human Behavior*, V. S. Ramachandran (ed.). San Diego, CA: Academic Press, pp. 436–443.

Moss, C. Scott (1972). *Recovery with Aphasia: The Aftermath of My Stroke*. Champaign, IL: University of Illinois Press.

Mr. Robot. "eps2.2_init_1.asec." Season 2, Episode 4. Directed by Sam Esmail. Written by Sam Esmail. USA Network, July 27, 2016.

Nagel, Thomas (1970). *The Possibility of Altruism*. Princeton: Princeton University Press.

Neale, Stephen (1999). "Coloring and Composition." In *Philosophy and Linguistics*, K. Murasugi and R. Stainton (eds.). Boulder, CO: Westview Press, pp. 35–82.

Norton, David F. and Mary J. Norton, eds. (2000). *David Hume: A Treatise of Human Nature*. New York: Oxford University Press.

Olinder, Ragnar Francén (2011). "Svavarsdóttir's Burden." *Philosophia*, Online First (November), pp. 1–13.

Penny, Laurie (2016). "I'm With The Banned." *Welcome to the Scream Room*. Medium Corporation, July 21, 2016. URL = <https://medium.com/welcome-to-the-scream-room/im-with-the-banned-8d1b6e0b2932#.b0aurvmnb>.

Piaget, Jean (1959). *The Language and the Thought of the Child*. New York: Routledge.

Prinz, Jesse (2000). "A Neurofunctional Theory of Visual Consciousness." *Consciousness and Cognition* Vol. 9, pp. 243–59.

Prinz, Jesse (2006). "The Emotional Basis of Moral Judgments." *Philosophical Explorations* Vol. 9, pp. 29–43.

Prinz, Jesse (2007). "All Consciousness is Perceptual." In *Contemporary Debates in the Philosophy of Mind*, J. Cohen and B. McLaughlin (eds.). Oxford: Blackwell, pp. 335–357.

Railton, Peter (2006). "Moral Factualism." In *Contemporary Debates in Moral Theory*, J. Dreier (ed.). Malden, MA: Blackwell Publishing Ltd, pp. 201–219.

Raphael, D. D. and A. L. Macfie, eds. (1982). *Adam Smith: The Theory of Moral Sentiments*. Indianapolis, IN: Liberty Fund.

Ridge, Michael (2006). "Sincerity and Expressivism." *Philosophical Studies* Vol. 131, pp. 487–510.

Ridge, Michael (2014). *Impassioned Belief*. New York: Oxford University Press.

Robson, Jon (2015). "Norms of Belief and Norms of Assertion in Aesthetics." *Philosopher's Imprint*, Vol. 15: <http://www.philosophersimprint.org/015006/>.

Rogers, Ben (1999). *A. J. Ayer: A Life*. New York: Grove Press.

Rojszczak, Artur and Barry Smith (2003). "Theories of Judgement." In *The Cambridge History of Philosophy 1870–1945*, T. Baldwin (ed.). New York: Cambridge University Press, pp. 157–173.

van Roojen, Mark (2009). "Moral Cognitivism vs. Non-Cognitivism." *The Stanford Encyclopedia of Philosophy* (Spring 2011 Edition), Edward N. Zalta (ed.). URL = <http://plato.stanford.edu/archives/spr2011/entries/moral-cognitivism/>.

van Roojen, Mark (2010). "Moral Rationalism and Rational Amoralism." *Ethics* Vol. 120 (April), pp. 495–525.

Rowling, J. K. (1999). *The Prisoner of Azkaban*. New York: Scholastic Press.

Sander, Thorsten (2016). "The Case of the Dissapearing Semi-Colon: Expressive Assertivism and the Embedding Problem." *Philosophia* Online First, pp. 1–21.

Sannino, Annalisa (2008). "Experiencing Conversations: Bridging the Gap between Discourse and Activity." *Journal for the Theory of Social Behaviour* Vol. 38, pp. 267–291.

Scanlon, T. M. (1998). *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

Schiffer, Stephen R. (1972). *Meaning*. New York: Oxford University Press.

Schroeder, Mark (2009). "Hybrid Expressivism: Virtues and Vices." *Ethics* Vol. 119, pp. 257–309.

Schroeder, Mark (2014). "The Truth in Hybrid Semantics." In *Having It Both Ways: Hybrid Theories and Modern Metaethics*, G. Fletcher and M. Ridge (eds.). New York: Oxford University Press, pp. 273–293.

Searle, John (1969). *Speech Acts*. New York: Cambridge University Press.

Sellars, Wilfred (1969). "Language as Thought and as Communication." *Philosophy and Phenomenological Research* Vol. 29, pp. 506–527.

Sennet, Adam and David Copp (2014). "What Kind of a Mistake is It to Use a Slur?" *Philosophical Studies* Vol. 172, pp. 1079–1104

Shafer-Landau, Russ (2003). *Moral Realism: A Defense*. New York: Oxford University Press.

Slingerland, Edward G., trans. (2005). "Kongzi (Confucius): 'The Analects.'" In *Readings in Classical Chinese Philosophy*, P. J. Ivanhoe and B. W. Van Norden (eds.). Indianapolis: IN, Hackett Publishing Company, pp. 1–57.

Smith, Michael (1994). *The Moral Problem*. Malden, MA: Blackwell Publishing Ltd.

Sneddon, Andrew (2009). "Alternative Motivation: A New Challenge to Moral Judgment Internalism." *Philosophical Explorations* Vol. 12, pp. 41–53.

Spoken Werd (2016). "Ted Cruz is a Pussy - Trump Rally Goer Calls Cruz Pussy." Online video clip, YouTube. YouTube, 8 February 2016. Web, 10 August 2016. URL = <https://www.
youtube.com/watch?v=XX4MO7YHqeg>.

Stevenson, Charles (1937). "The Emotive Meaning of Ethical Terms." *Mind* Vol. 46, pp. 14–31.

Strandberg, Caj (2011a). "The Pragmatics of Moral Motivation." *Journal of Ethics* Vol. 14, 341–369.

Strandberg, Caj (2012). "A Dual Aspect Account of Moral Language." *Philosophy and Phenomenological Research* Vol. 84, pp. 87–122.

Strandberg, Caj (2013). "An Internalist Dilemma–and an Externalist Solution." *Journal of Moral Philosophy* Vol. 10, pp. 25–51.

Strawson, P. F. (1964). "Intention and Convention in Speech Acts," *The Philosophical Review* Vol. 73, pp. 439–60.

Strawson, P. F. (1970). *Meaning and Truth: An Inaugural Lecture Delivered Before the University of Oxford on 5 November 1969*. Oxford: Oxford University Press.

Svavarsdóttir, Sigrún (1999). "Moral Cognitivism and Motivation." *Philosophical Review* Vol. 108, 161–219.

Sytsma, Sharon (1995). "Ethical Internalism and Moral Indifference." *The Journal of Value Inquiry* Vol. 29, pp. 193–201.

Tappan, Mark B. (1990). "Hermeneutics and Moral Development: Interpreting Narrative Representations of Moral Experience." *Developmental Review* Vol. 10, pp. 239–265.

Tappan, Mark B. (1997). "Language, Culture, and Moral Development: A Vygotskian Perspective." *Developmental Review* Vol. 17, pp. 78–100.

Tappan, Mark B. (2006a). "Mediated Moralities: Sociocultural Approaches to Moral Development." In *Handbook of Moral Development*, M. Killen and J. Smetana (eds.). Mahwah, NJ: Lawrence Erlbaum Associates, pp. 351–374.

Tappan, Mark B. (2006b). "Moral functioning as mediated action." *Journal of Moral Education* Vol. 35, pp. 1–18.

Thomson, Judith Jarvis (1996). "Emotivism." In *Moral Relativism and Moral Objectivity*, G. Harman and J. Thomson (eds.). Malden, MA: Blackwell Publishing Ltd., pp. 95–124.

Unwin, Nicholas (1999). "Quasi-Realism, Negation, and the Frege Geach Problem." *The Philosophical Quarterly* Vol. 49, pp. 337–352.

Vicente, Agustin and Fernando Martinez Manrique (2011). "Inner Speech: Nature and Functions." *Philosophy Compass* Vol. 6, pp. 209–219.

Vygotsky, Lev (1986). *Thought and Language*. Cambridge, MA: The MIT Press.

Westermarck, Edward (1912). *The Origin and Development of the Moral Ideas*. London: Macmillon and Co.

Wiley, Norbert (2006a). "Inner Speech as a Language: A Saussurean Inquiry." *Journal for the Theory of Social Behavior* Vol. 36, pp. 319–341.

Wiley, Norbert (2006b). "Peirce and the Founding of American Sociology." *Journal of Classical Sociology* Vol. 6, pp. 23–50.

Williams, Bernard (1982). *Moral Luck*. New York: Cambridge University Press.

Williams, Bernard (1985). *Ethics and the Limits of Philosophy*. Cambridge, MA: Harvard University Press.

Williamson, Timothy (1996). "Knowing and Asserting." *The Philosophical Review* Vol. 105, pp. 489–523.

Winsler, Adam (2009). "Still Talking to Ourselves after All These Years: A Review of the Current Research on Private Speech." *In Private Speech, Executive Functioning, and the Development of Verbal Self-Regulation*, A. Winsler, C. Fernyhough, and I. Montero (eds.). New York: Cambridge University Press, pp. 3–41.

Winsler, Adam, Charles Fernyhough and Ignacio Montero (2009). "Introduction." In *Private Speech, Executive Functioning, and the Development of Verbal Self-Regulation*, A. Winsler, C. Fernyhough, and I. Montero (eds.). New York: Cambridge University Press, pp. xi–xvii.

Young-Bruehl, Elizabeth (1982). "Reflections on Hannah Arendt's *The Life of the Mind*." *Political Theory* Vol. 10, pp. 277–305.