ROBERT B. BRANDOM

Between Saying & Doing

TOWARDS AN ANALYTIC PRAGMATISM

Between Saying and Doing:
Towards an Analytic Pragmatism

This page intentionally left blank

Between Saying and Doing:

Towards an Analytic Pragmatism

Robert B. Brandom

OXFORD
UNIVERSITY PRESS

In fond and loving memory, this work is dedicated to a great man, a great teacher, and a great friend,

Richard Rorty

for whom I am still rewriting my dissertation.

This page intentionally left blank

Contents Summary

This page intentionally left blank

Contents

Preface

We are familiar with the idea that a complete account of *semantics* (the study of the *meanings* expressed by different vocabularies) can require attention also to *pragmatics* (the study of the *use* of vocabularies). The paradigm is *indexical* and *demonstrative* expressions. Determining the referent of tokenings of types such as 'I', 'here', 'now', 'this', and 'that' depends on features of the use of those very tokenings: who uttered them, where, when, and in what surrounding context. A principal contention of this work is that the class of vocabularies whose meanings cannot be specified without attending to features of use is much larger, and the sorts of dependence of meaning on use much more varied, than is suggested by focusing on this central kind of example. I argue that alethic modal vocabulary, deontic normative vocabulary, and even classical logical vocabulary are also of this general kind. They differ from indexical and demonstrative expressions in that what is essential to understanding their semantics is their relation to the use of *other* vocabularies (non-modal, non-normative, non-logical ones), not to the pragmatic features of their own tokenings. One of the reasons this kind of semantic relation has been overlooked is an overemphasis on semantic *compositionality*. In the fifth of these lectures I offer some specific technical results aimed at clarifying some of the issues in the vicinity in a way that may lead to a more balanced assessment.

Modal, normative, and logical vocabularies are all ones that have been of central, perennial interest to philosophers of language over the past century or so. But this distinctive and essential feature of their semantics—as I will put it, the constellation of *pragmatically mediated* semantic relations they stand in to other vocabularies—has not been studied, or even (with some notable exceptions) much noticed by the tradition. The culmination of my argument in this book is that focusing on relations of this kind makes visible crucial features of what is expressed by the use of *semantic*, and more broadly, *intentional* vocabulary. (The title of the sixth and final lecture is "Intentionality as a Pragmatically Mediated Semantic Relation.") Further, this pragmatically mediated semantic feature of semantic and intentional

vocabularies is intimately related to the pragmatically mediated semantic relations in which alethic modal and deontic normative vocabularies stand to each other.

Along the way to this result—as a necessary means of formulating, clarifying, and justifying it—I introduce, and develop through application to other examples, a general method of formally representing relations between *meaning* and *use* (in the broad way in which I will use the terms, between the topics of *semantics* and *pragmatics*). I call it "meaning-use analysis." Analyses of this kind are codified and expressed in their most perspicuous form in meaning-use *diagrams*, which the reader will find peppered throughout the book. Each such diagram presents one or more of the infinite number of pragmatically mediated semantic relations among vocabularies, whose representations are recursively generated by the methods of meaning-use analysis. By way of sharpening and making this conceptual apparatus more definite, I consider along the way a number of further applications of it to a variegated budget of examples: computational linguistics, artificial intelligence, even indexical vocabulary itself. In this same spirit, I also use it to introduce a new kind of formal semantics for alethic modal and classical logical vocabularies that brings to light hitherto invisible relations between them.

The result is an approach to the philosophy of language that in my subtitle I call an "analytic pragmatism." It is pragmatism pursued in an analytic spirit. By calling it "pragmatism" I mean a view inspired by insights of the later Wittgenstein, which situates concern with the *meanings* of expressions in the broader context of concern with proprieties governing their *use*. It counsels us to start our thinking about the meanings expressed by various vocabularies by thinking about discursive practices, about the abilities whose exercise constitutes using the vocabularies to express those meanings. Pursuing those pragmatist ideas in an analytic spirit is rejecting the anti-theoretical, anti-systematic conclusions that are often drawn from them. Instead, we can, the claim is, think about the relations between meaning and use every bit as rigorously and systematically as it has proven possible to think about the sorts of relations between meanings that are codified and explored in classical formal semantics, for instance as developed within the dominant Tarskian model-theoretic and possible worlds frameworks.

★

This work presents my first big philosophy of language project since the one reported in *Making It Explicit*.[1] As I indicate in more detail in the Afterword, these are different projects. Although they have some topics in common (for instance, both are concerned with the relations between semantics and pragmatics, and with the nature of logic and its role in philosophy), this book is not a further working-out of the theory presented in that earlier one. The two are largely orthogonal enterprises. I think they are broadly compatible, but I have not worried overmuch about reconciling, or even relating, them in a compare-and-contrast sort of way (and devote none of the present volume to doing any of that).

The genesis and development of the two works is also as different as it well could be. Both are ambitious, and each is in its own way systematic. But *Making It Explicit* took shape very slowly. Working out the details so as to fill in each part of the initial plan, adjusting the remainder as necessary so as to balance the sometimes conflicting demands of both remaining true to the original animating vision and taking suitable account of the lessons learned along the way, took eighteen years to produce a stable product I could call finished. (I've been at my Hegel project even longer than that.) I was determined throughout to take as long as it took. *Ars*, indeed, *longa*, however *brevis* the *vita*. By contrast, this project came from the beginning with a deadline: a date, fixed in advance, on which I would have to stand up and present the material to a demanding audience. In the spring of 2002 I received from the Oxford Philosophy Faculty the kind invitation to deliver the John Locke Lectures there. Their initial suggestion was that I do so in 2004. I felt strongly that that was not enough time to prepare new work worthy of the occasion, and suggested the following year. It turned out that a previous, equally conscientious and foresighted invitee (Ernie Sosa), had had the same sort of resistance to a short lead time, and so was already scheduled for 2005. So we settled on 2006. (Four years is today a not atypical length of time for someone to spend writing a philosophy PhD dissertation—a work of about the same length I would be aiming for.)

---

[1] Harvard University Press, 1994.

My initial response to the invitation was blind panic. Although the Oxford people professed politely to be willing to listen to anything I might have to say, it was clear to me that I had *not* been invited to talk about Hegel—the topic then occupying the center of my attention. Further, under what they charmingly called the ''McDowell rule,'' Oxford University Press, who financially support the lectures, now insist that one explicitly agree in advance to let them have first consideration of any book arising out of the material presented in those lectures. (Prior practice had been to leave this expectation implicit. McDowell refused to comply, sending his deep, path-breaking book *Mind and World* instead to Harvard, in part because of OUP's having let Gareth Evans's work go out of print.) I had no objection to such a distinguished venue. But my long, long-gestating Hegel book has long been promised to my long-suffering editor at the Harvard University Press, Lindsay Waters. So the Hegel was out. I felt strongly that this invitation should be understood as a challenge to present the best philosophy of which one is capable. Rehashing, or even just extending, the views I had already presented in excruciating detail in *Making It Explicit* and subsequent works would be failing to rise to the occasion. I have a good deal of material on Frege, which I had been working up over a period of time, with an eye to an eventual monograph. But even my most optimistic view of what it could aspire to achieve would not put it in the same class as the contribution made by Macbeth's fabulous book *Frege's Logic*,[2] whose manuscript I had recently been privileged to read. I would have been glad to present *her* work, had I written it, but was not going to set myself up to suffer the comparison.

Then it occurred to me that I did have a plan for a three-article set that would develop the ideas I had sketched earlier in a piece called ''Modality, Normativity, and Intentionality.''[3] There I had stated and explored some of the consequences of accepting what I called the ''Kant-Sellars theses'' about alethic modal and deontic normative vocabularies. These are the claims, roughly, that anyone who knows how to use ordinary, non-modal, non-normative vocabulary already knows how to do everything she needs to know how to do to deploy also modal and normative vocabulary. In this sense, those kinds of vocabulary have the expressive role of making

---

[2] Danielle Macbeth, Harvard University Press, 2005.

[3] *Philosophy and Phenomenological Research* 63, no. 3 (November 2001), 587–609, with a comment by Gideon Rosen.

explicit something that is implicit already in the use of ordinary descriptive vocabulary. But I had not offered *arguments* that those claims *are* true, nor thought very much about what *kind* of claim they are. I proposed to do that in the first of the three elements of the triptych. In the next, I hoped to make good on a long-term project of using the notion of <u>material incompatibility</u> as the semantic primitive of a new sort of formal semantics—one that would represent the content of claims (including those formulated using descriptive, modal, and logical vocabulary) in terms of the set of claims materially incompatible with them. The *use* of this ultimately *modal* semantic primitive could in turn be introduced in terms of the *normative* statuses of commitment and entitlement, according to the principle that to treat claims as materially incompatible is to take *commitment* to one to preclude *entitlement* to the other. I thought this idea could be developed into a unified setting that would permit the representation of the contents of alethic modal claims, about what is possible and what is necessary, at exactly the same level and in the same terms as the contents of ordinary logical claims, as well as non-logical ones. But there were substantial technical challenges to be overcome, which at the time I had as yet made little progress on, despite fiddling with them over a period of years. Finally, to complete the triad, I wanted to exploit the relations between what is expressed by deontic normative vocabulary (paradigmatically 'committed' and 'entitled') and alethic modal vocabulary ('necessary' and 'possible') that were revealed by putting together the Kant-Sellars theses with the way a semantics for *modal* vocabulary could be elaborated from what is expressed by *normative* vocabulary. In those terms I thought I could say something new and interesting about the intentional nexus between knowing and acting *subjects*, who are obliged to reject or resolve incompatible *commitments* and *objects*, which are individuated in part by the impossibility of their exhibiting incompatible *properties*. (I had already had the idea that something like these two senses of ''determinate negation''—his term for ''material incompatibility''—lies behind some of Hegel's darker, but also deeper, claims about the relations between the subjective and the objective.[4]) It would be worthwhile, I thought, to try to motivate and elaborate this idea wholly in contemporary terms, without having to say anything about Hegel.

---

[4] See chapter 6 of my *Tales of the Mighty Dead: Historical Essays in the Metaphysics of Intentionality* (Harvard University Press, 2002).

This was an idea for three articles or lectures, and the Locke Lectures were to be six. But I thought it likely that there was enough material there to make up six meaty lectures, if I just thought about it all hard enough. It was on that basis that the panic subsided (it took about two weeks), and I felt I could responsibly accept the invitation. In the presentation that eventually emerged, the three pieces I had originally conceived correspond pretty well to the last three of the six lectures. The surprising and (as it seems to me) suggestive story about intentionality and its relation to what is expressed by alethic modal and deontic normative vocabularies that unfolds in the last three lectures ended up prefaced by a somewhat more abstract methodological story told in the first three, which develops some of the lessons about intentionality appealed to in the final lecture. The big challenge—since the whole thing would not work without it—was getting the technical results needed to make the incompatibility semantics work. Through great good fortune (though the timing was not just a bizarre coincidence), I was able to spend the academic year 2002–2003 as a fellow at the Center for Advanced Study in the Behavioral Sciences. (This is where Quine wrote *Word and Object*, and Rawls wrote *A Theory of Justice*—but no pressure!) Though I was not aware of it in advance, it turns out that my fellowship there was financially supported by the Andrew W. Mellon Foundation—just the first of their contributions to this project. Able to devote myself full-time to philosophical work, I made good progress on my Hegel project: producing drafts of the three chapters I was then in a position actually to write. But I devoted substantial pie-in-the-sky, staring-off-into-space, let's-try-anything-that-might-work time to the incompatibility semantics. By the end of that golden period, I had made the basic conceptual advances (along with writing some useful computer programs) that I was confident would eventually result in a successful up-and-running product. (It turned out to be a bumpier ride to that destination than I had thought it would be.)

Upon my return from Stanford, I began to think about the overall project again. It was at this point that I realized that the Kant-Sellars theses asserted a distinctive kind of pragmatically mediated semantic relation between modal and normative vocabularies, on the one hand, and ordinary descriptive vocabulary, on the other. Trying to characterize the genus of which this new sort of semantic relation was a species led to the first meaning-use diagrams, and then to the more general idea of meaning-use

analysis. Looking for an index example, where the results of a non-trivial meaning-use relation were not philosophically controversial (as all my other proposed examples were bound to be), I realized that I had stumbled on a new kind of metavocabulary for discussing the relations between automata and vocabularies that are studied in the relatively clear-cut syntactic arena by computational linguistics. This provided the example I exploit in the first lecture, to introduce the more general metaconceptual apparatus.

I had for many years taught an undergraduate lecture course at Pitt (originally conceived by John Haugeland, based on his fine textbook *Artificial Intelligence: The Very Idea*[5]) that used current debates about the possibility of artificial intelligence to introduce more general themes in the philosophy of mind. So I was familiar with that literature, and had had the opportunity to think through some of the issues. I soon realized that the automaton-theoretic issues I had been considering at the level of syntax would generalize nicely to the semantic level, and that, when they did, the result would be a substantially new way of thinking about the central issues of artificial intelligence—one that shows debates about the essentially symbolic character of thought to be parochial and peripheral. This line of thought led to the formulation and (skeptical) discussion in the third lecture of a *pragmatist* (as opposed to the classical intellectualist) version of the thesis of artificial intelligence: the claim that the capacity to engage in some autonomous discursive practice (a language-game one could play though one played no other) could be algorithmically decomposed into abilities each of which can, in principle, be exhibited by non-discursive systems.

I taught all this new material in my graduate seminar in the winter of 2005. So that this idiosyncratic stuff would have some context, and the students would learn at least something of value even if the rest of it fell apart, I started off by talking about the history of analytic philosophy from Russell to Quine, as a foil to the pragmatist line I would be taking. In the course of telling that story, I realized that in the systematic form in which I could now tell my pragmatist story, that form need not at all be thought of as an alternative to the analytic tradition. Rather, once the classical project of philosophical analysis is described broadly enough to

---

[5] MIT Press, 1989.

encompass most of its principal variants—including at least Russell and Moore, the *Tractatus*, the Cambridge analysts of the 1920s, the Vienna Circle, Ayer, Goodman, and Quine—what I was doing could be seen at least as much as a continuation and extension of that project, bringing substantial new expressive resources to bear on its behalf. That is how I came to introduce the project in the first lecture. At that point, almost exactly a year before I was scheduled to deliver the lectures, I sat down to produce the actual text.

By happy accident, at the end of 2003 I had received a letter telling me I had been awarded a Distinguished Achievement in the Humanities prize from the Andrew W. Mellon Foundation. This astonishingly generous award provides one and a half million dollars to enable its recipient to pursue his or her research in ways not otherwise feasible. Though I have found many other (I trust) worthwhile endeavors to which to apply these funds (for instance, cataloguing and making available to the scholarly public the papers of Wilfrid Sellars), the project reported in these pages has been a principal beneficiary and product of the extraordinary opportunity they afford. I indicated earlier that I knew from the beginning that the technical adequacy of the new modal incompatibility semantics I propose was going to be a critical criterion of adequacy determining whether the story could take the shape I hoped it could. The Mellon money made it possible to assemble a research team to help work out and exploit the basic ideas. My colleague Ken Manders, my student Jukka Keranen, and my research assistants Jeremy Heis and Alp Aker, all made significant contributions to moving the work forward, and to furthering our understanding of the, in many ways, very unusual formal setting of this semantics. (For instance, as detailed in Lecture 5 and its long technical appendix, this is a *holistic* semantics that is nonetheless fully *recursive*. At the level of non-logical sentences, it corresponds to a multivalued logic in which the number of multivalues varies depending on how many atomic sentences there are in the language.) The one among these who made the greatest contribution is Alp Aker. He is responsible for almost all of the metatheoretic results reported in the appendix to Lecture 5. Without his ingenuity, insight, and hard work the basic ideas would not have been developed nearly as far nor as well as they are here. Nonetheless, it is early days, and what is offered here is offered in the spirit of a report on work in progress.

The lectures that appear here are essentially as I delivered them at Oxford in Trinity Term of 2006. I have resolutely resisted the temptation to write the long book that is struggling to get out of this short one. Brevity (in art, if not in life) has its virtues. The exception is the two appendices, which were already available when the lectures were given. They provide detail necessary to back up claims made in the lectures proper. Only the first half of the argument concerning indexical vocabulary is presented in the first lecture. The argument would be seriously incomplete without the material in the appendix to Lecture 2, which simply could not be shoehorned into the lectures themselves without pushing out something even more important to the whole project. (Material developed only in the second lecture is needed to fill in the argument about indexical vocabulary gestured to in the first.) The long technical appendix to Lecture 5 provides the definitions and proofs that articulate incompatibility semantics for modal and classical logic. They are the cash for the promissory notes offered by my descriptions of those results in the lecture. As I remark there, in this context, proof is the word made flesh. Those who are willing to take my word for it can skip the dreary details. But anyone who wants to understand how the system really works has to have them all available.

The John Locke Lectures are delivered once a week over the course of a term at Oxford, to an audience whose composition inevitably varies with the exigencies of individual schedules. I was both fortunate in and flattered by the relatively large numbers who continued to turn out. (I'm sure that the exhortations of my gracious principal host, Tim Williamson, helped secure that happy result.) This is a point about which I had been somewhat apprehensive. My friend and colleague John McDowell got great audiences for his 1991 Locke Lectures (as well he should have, given the material he presented), but when Sellars had done his in 1965, the precipitous drop-off in attendance, ending with an audience for the last three of five or six souls scattered through a cavernous auditorium, is something he was bitter about literally until his dying day. (When I visited him in the hospital just before his death, he cited the disparity between the reception of his Locke Lectures and that accorded his University of London lectures on "Empiricism and the Philosophy of Mind" just nine years earlier, as one of the great disappointments of his professional life.) The question periods, and other less formal discussions, including weekly meetings with graduate students, were often helpful to me.

The most intense and thought-provoking conversations about this material, though, took place in Prague, in April of 2007. I had been keen to make these ideas vulnerable to the criticism of a wider philosophical audience. The largesse of the Mellon Foundation and the good offices of my friends Jaroslav Peregrin of the Charles XII University and the Czech Academy of Sciences, and Pirmin Stekeler of the University of Leipzig, made it possible. The result was a week-long meeting, attended by more than two hundred philosophers from twenty-seven countries, including a host of specially invited people whose opinions I particularly value. The lecture portion of the meeting was preceded (at Jarda Peregrin's suggestion) by a three-day ''School of Inferentialism,'' in which people intimately familiar with it presented papers and presided over informational sessions devoted to various aspects of my prior work, as background for the new material everyone would hear later. (Michael Williams, James Conant, Björn Ramberg, Mark Lance, Michael Kremer, and Paul Horwich were the presenters.) Each of my six lectures was followed by a paper discussing it. In order of presentation, these were: John McDowell, John MacFarlane, Pirmin Stekeler, Huw Price, Jaroslav Peregrin, and Sebastian Rödl. I was particularly touched that my Doktorvater, Richard Rorty, having known for some time that he was mortally ill, and with the end near (he died on 8 June 2007), made the extraordinary gesture, appreciated by all, to come and to participate. These lectures are dedicated to him—my generous teacher, friend, best critic, and the first reader of these lectures, as he has been of most of what I have written over the past thirty-five years.

In July 2007, I gave all six lectures once more, this time to an impressively sophisticated, varied, and engaged audience at the University of Buenos Aires. Although I have made some relatively minor emendations in response to the comments elicited at all these events, I decided in the end that the material was best served by being presented in something very close to its original form. I have confined myself to appending to the lectures an afterword, in which I respond to some of the larger-scale worries—particularly those expressed by John McDowell and Sebastian Rödl (though similar notes have been sounded in the comments of others). The Afterword offers such arguments as occur to me that the analytic–algebraic form in which I have presented my thoughts should be understood not as fetters binding the radical, reconstructive spirit of Rorty and the experimental,

exploratory spirit of David Lewis, but rather as a vocabulary in which to express, apply, and develop those thoughts, consonant with both of those spirits.

Bob Brandom
August 2007

This page intentionally left blank

Acknowledgements

List of Abbreviations

AI	artificial intelligence
ADP	autonomous discursive practice
FSA	finite-state automaton
FSTA	finite-state transducing automaton
LSA	laughing Santa automaton
MUA	meaning-use analysis
MUD	meaning-use diagram
MUR	meaning-use relation
PDA	push-down automaton
PV	practice-vocabulary
SSTA	single-state transducing automaton
VP	vocabulary-practice

Between saying and doing, many a pair of shoes is worn out.
Italian proverb

This page intentionally left blank

1

Extending the Project of Analysis

My aim in these lectures is to present a new way of thinking about language, specifically about the relations between meaning and use, or between what is said and the activity of saying it. To that end, I will introduce a new metatheoretic conceptual apparatus, and develop it through applications to a number of sorts of locution that have, properly, been the focus of intense philosophical interest: logical and semantic vocabulary, indexical vocabulary, modal, normative, and intentional vocabularies. The concerns that animate this enterprise arise from a way of thinking about the nature of the general project pursued by analytic philosophy over the past century or so, and about its epic confrontation with Wittgensteinian pragmatism. Justifying that rendering of the tradition would take me far afield, but it will be well to begin with at least a sketch of that motivating picture.

I   The classical project of analysis

I think of analytic philosophy as having at its center a concern with semantic relations between what I will call 'vocabularies'. Its characteristic form of question is whether, and in what way, one can make sense of the meanings expressed by *one* kind of locution in terms of the meanings expressed by *another* kind of locution. So, for instance, two early paradigmatic projects were to show that everything expressible in the vocabulary of number theory, and again, everything expressible using definite descriptions, is expressible already in the vocabulary of first-order quantificational logic with identity.

The nature of the key kind of semantic relation between vocabularies has been variously characterized during the history of analytic philosophy: as analysis, definition, paraphrase, translation, reduction of different sorts, truth-making, and various kinds of supervenience—to name just a few contenders. In each case, however, it is characteristic of classical analytic philosophy that *logical* vocabulary is accorded a privileged role in specifying these semantic relations. It has always been taken at least to be *licit* to appeal to logical vocabulary in elaborating the relation between *analysandum* and *analysans*—target vocabulary and base vocabulary. I will refer to this aspect of the analytic project as its commitment to '*semantic logicism*'.[1]

If we ask *which* were the vocabulary-kinds whose semantic relations it was thought to be important to investigate during this period, at least two *core programs* of classical analytic philosophy show up: *empiricism* and *naturalism*. These venerable modern philosophical traditions in epistemology and ontology respectively were transformed in the twentieth century, first by being transposed into a *semantic* key, and second by the application of the newly available *logical* vocabulary to the self-consciously semantic programs they then became.

As *base* vocabularies, different species of *empiricism* appealed to phenomenal vocabulary, expressing how things appear, or to secondary-quality vocabulary, or, less demandingly, to observational vocabulary. Typical *target* vocabularies include objective vocabulary formulating claims about how things actually are (as opposed to how they merely appear), primary-quality vocabulary, theoretical vocabulary, and modal, normative, semantic, and intentional vocabularies. The generic challenge is to show how what is expressed by the use of such target vocabularies can be reconstructed from what is expressed by the base vocabulary, when it is elaborated by the use of logical vocabulary.

As base vocabularies, different species of *naturalism* appealed to the vocabulary of fundamental physics, or to the vocabulary of the natural sciences (including the special sciences) more generally, or just to objective descriptive vocabulary, even when not regimented by incorporation into explicit scientific theories. Typical targets include normative, semantic, and intentional vocabularies.

---

[1] In this usage, the logicism about mathematics characteristic of Frege's *Grundgesetze* and Russell and Whitehead's *Principia* is semantic logicism about the relations between mathematical and logical vocabularies.

2 The pragmatist challenge

What I want to call the "classical project of analysis," then, aims to exhibit the meanings expressed by various target vocabularies as intelligible by means of the logical elaboration of the meanings expressed by base vocabularies thought to be privileged in some important respects—epistemological, ontological, or semantic—relative to those others. This enterprise is visible in its purest form in what I have called the "core programs" of empiricism and naturalism, in their various forms. In my view, the most significant conceptual development in this tradition—the biggest thing that ever happened to it—is the *pragmatist challenge* to it that was mounted during the middle years of the twentieth century. Generically, this movement of thought amounts to a displacement from the center of philosophical attention of the notion of <u>meaning</u> in favor of that of <u>use</u>: in suitably broad senses of those terms, replacing concern with *semantics* by concern with *pragmatics*. The towering figure behind this conceptual sea-change is, of course, Wittgenstein. In characterizing it, however, it will be useful to approach his radical and comprehensive critique by means of some more local, semantically corrosive argumentative appeals to the *practices of deploying* various vocabularies rather than the *meanings* they express.

Wilfrid Sellars (one of my particular heroes) criticizes the empiricist core program of the classical project of analysis on the basis of what one must *do* in order to *use* various vocabularies, and so to count as *saying* or *thinking* various kinds of things. He argues that *none* of the various candidates for empiricist base vocabularies is practically *autonomous*, that is, could be deployed in a language-game one played though one played no other. For instance, no discursive practice can consist entirely of making non-inferential observation reports, for such reliably differentially elicited responses qualify as *conceptually* contentful or *cognitively* significant only insofar as they can serve as *premises* from which it is appropriate to draw *conclusions*, that is, as *reasons* for other judgments. Drawing such conclusions is applying concepts *inferentially*—that is, precisely *not* making non-inferential *observational* use of them.[2]

<hr>

[2] This argument occupies roughly the first half of his classic "Empiricism and the Philosophy of Mind" (1956; reprinted by Harvard University Press, 1997). His critique of the phenomenalist version of empiricism can be found in "Phenomenalism," in his collection *Science, Perception, and Reality*

Quine offers an even broader pragmatist objection, not only to the empiricist program, but to essential aspects of the whole analytic semantic project, for he attacks the very notion of <u>meaning</u> it presupposes. Quine is what I have elsewhere called a "methodological" pragmatist.[3] That is, he takes it that the whole point of a theory of meaning is to explain, codify, or illuminate features of the *use* of linguistic expressions. He, like Dummett, endorses the analogy: *meaning* is to *use* as *theory* is to *observation*. And he argues that postulating meanings associated with bits of vocabulary yields a *bad* theory of discursive practice.

If there were such things as meanings that determine how it would be correct to use our expressions, then their meanings would at least have to determine the inferential roles of those expressions: what follows from applying them, what applying them rules out, what is good evidence for or against doing so. But what follows from what depends on what else is true—on laws of nature and obscure contingent facts—that is, on what claims can serve as auxiliary hypotheses or collateral premises in those inferences. If we look at what practical abilities are required to deploy various bits of vocabulary—at what one has to be able to *do* in order to count as *saying* something with them—we do not find any special set of these whose practical significance can be understood as *pragmatically* distinctive of *semantically* necessary or sufficient conditions.

Quine thought one could save at least the naturalist program by retreating semantically to the level of reference and truth-conditions. James and Dewey appeal to the same sort of methodological pragmatism in support of more sweeping sorts of semantic revisionism—pursuing programs that Rorty, for instance, argues should be understood as more rejectionist than properly revisionist. And under the banner "Don't look to the meaning, look to the use," Wittgenstein further radicalizes the pragmatist critique of semantics. Pointing out, to begin with, that one cannot assume that uses of singular terms have the job of picking out objects, nor that declarative sentences are in the business of stating facts, he goes on to deny, in effect, that such uses even form a privileged center on the basis of which

one can understand more peripheral ones. ("Language," he says, "has no downtown.")

I take it that Wittgenstein also understands the home language-game of the concept of <u>meaning</u> to be explanation of how expressions are correctly *used*. And he is profoundly skeptical about the utility or applicability of the model of postulation, explanation, and theoretical systematization in the case of discursive practices—about the possibility of systematically *deriving* aspects of correct use from assigned meanings. Seen from this perspective, the idea of the classical project of analysis is to codify, using logical vocabulary, the meanings expressed by one vocabulary—from which we are to derive proprieties of its use—from the meanings expressed by some *other* vocabulary—from which we can derive proprieties of *its* use. One of his ideas, I think, is that this enterprise makes sense only if we think of the uses as species of a genus—of them all being the same general *kind* of use, say describing, stating facts, or representing states of affairs. This may seem plausible if we focus on a very restricted set of uses—just as, in the case of tools, we might be impressed to notice that nails and hammer, screws and screwdriver, glue and brush all have the function of attaching more-or-less flat things to one another. So we can think of declarative sentences as stating empirical, physical, normative, modal, and intentional facts, making *claims* about such states of affairs (even if we then find ourselves metaphysically puzzled about the nature of the fact-kinds to which we have thereby committed ourselves). But if we think of the uses as *very* different, if we think also about the carpenter's level, pencil, and tool-belt, if we think of linguistic practice as a *motley*, of uses as not coming in a simple, or systematic, or even determinate variety, then the very idea that there is such a thing as *meanings* that permit the codification of proprieties of quite disparate kinds of use—even with liberal use of logical elaboration of the meanings—becomes contentious and in need of justification both in general and in each particular case.

More specifically, Wittgenstein uses the image of "family resemblances" to urge that the *kinds* into which linguistic practices and the vocabularies caught up in them are functionally sorted—what belong together in boxes labeled 'game', 'name', 'description', 'assertion', 'observation', and so on—do not typically admit of specification in terms of underlying principles specifiable in other vocabularies, whether by genus and differentia(e) or any other kind of explicit rule or definition. It is easy to understand this line of

thought as entailing a straightforward denial of the possibility of semantic analysis in the classical sense.

I think that one thought underlying these observations about the unsystematic, unsurveyable variety of kinds of uses of expressions and about the uncodifiable character of those kinds concerns the essentially *dynamic* character of linguistic practice. I think Wittgenstein thinks that an absolutely fundamental discursive phenomenon is the way in which the abilities required to deploy one vocabulary can be practically *extended*, elaborated, or developed so as to constitute the ability to deploy some further vocabulary, or to deploy the old vocabulary in quite different ways. Many of his thought-experiments concern this sort of process of *pragmatic projection* of one practice into another. We are asked to imagine a community that uses proper names only for people, but then extends the practice to include rivers. There is no guarantee that interlocutors can master the extended practice, building on what they can already do. But if they can, then they will have changed the only ˢessenceˢ proper-name usage could be taken to have had.[4] In the old practice it always made sense to ask for the identity of the *mother* and *father* of the named item; in the new practice, that question is often senseless. Again, we are asked to imagine a community that talked about having gold or silver in one's teeth, and extends that practice to talk about having pain in one's teeth. If, as a matter of contingent fact, the practitioners can learn to use the expression 'in' in the new way, building on but adapting the old, they will have fundamentally changed the ˢmeaningˢ of 'in'. In the old practice it made sense to ask where the gold was *before* it was in one's tooth; in the new practice asking where the pain was before it was in the tooth can lead only to a distinctively *philosophical* kind of puzzlement.[5]

At every stage, what practical extensions of a given practice are possible for the practitioners can turn on features of their embodiment, lives, environment, and history that are contingent and wholly particular to them. And which of those developments actually took place, and in what

---

[4] Cf. Quine's remark in "Two Dogmas of Empiricism": "Meaning is what essence becomes when it is detached from the thing and attached to the word" (1953; reprinted by Harvard University Press, 2006).

[5] I am indebted for this way of thinking of Wittgenstein's point to Hans Julius Schneider's penetrating discussion in his *Phantasie und Kalkul: Über die Polarität von Handlung und Struktur in der Sprache* (Suhrkamp, 1992).

order, can turn on any obscure fact. The reason vocabulary-kinds resist specification by rules, principles, definitions, or meanings expressed in other vocabularies is that they are the current time-slices of processes of development of practices that have this dynamic character—and that is why the collection of uses that is the current cumulative and collective result of such developments-by-practical-projection is a motley.[6] If that is right, then any codification or theoretical systematization of the uses of those vocabulary-kinds by associating with them meanings that determine which uses are correct will, if at all successful, be successful only contingently, locally, and temporarily. Semantics on this view is an inherently Procrustean enterprise, which can proceed only by *theoretically* privileging some aspects of the use of a vocabulary that are not at all *practically* privileged, and spawning philosophical puzzlement about the intelligibility of the rest.[7] On this conception, the classical project of analysis is a disease that rests on a fundamental, if perennial, misunderstanding—one that can be removed or ameliorated only by heeding the advice to replace concern with *meaning* by concern with *use*. The recommended philosophical attitude to discursive practice is accordingly *descriptive particularism, theoretical quietism*, and *semantic pessimism*.

3 Extending the project of analysis: pragmatically mediated semantic relations

On this account, Wittgenstein is putting in place a picture of discursive meaningfulness or significance that is very different from that on which the classical project of analysis is predicated. In place of *semantics*, we are encouraged to do *pragmatics*—not in the sense of Kaplan and Stalnaker, which is really the semantics of token-reflexive expressions, nor again in

---

[6] A patient and detailed investigation of the mechanisms of this phenomenon in basic descriptive and scientific concepts, and an extended argument for its ubiquity, can be found in Mark Wilson's exciting and original *Wandering Significance* (Harvard University Press, 2006).

[7] I would be happy if those who dance with his texts find affinities here with Hegel's insistence that the metaconceptual categories of *Verstand* must be replaced by those of *Vernunft*. It is characteristic of his philosophical ambition that he draws the opposite of Wittgenstein's conclusions from an appreciation of the dynamics of conceptual development and its sensitivity to arbitrary contingent features of the practitioners, devoting himself to elaborating what he insists is the *logic* of such processes and the conceptual contents they shape.

the sense of Grice, which addresses conversational heuristics in terms that presuppose a prior, independent, classical semantics—but 'pragmatics' in the sense of the study of the *use* of expressions in virtue of which they are meaningful at all. To the formal, mathematically inspired tradition of Frege, Russell, Carnap, and Tarski, culminating in model-theoretic and possible-worlds semantics, is opposed an anthropological, natural-historical, social-practical inquiry aimed both at demystifying our discursive doings and at deflating philosophers' systematic and theoretical ambitions regarding them. I think that contemporary philosophers of language have tended to draw this opposition in the starkest possible terms, treating these approaches as mutually exclusive, hence as requiring that a choice be made between them, thereby marking out a substantial sociological faultline in the discipline. Those who are moved by the pragmatist picture generally seem to accept the particularist, quietist conclusions Wittgenstein seems to have drawn from it. And those committed to some version of the project of semantic analysis have often felt obliged to deny the significance of pragmatics in this sense, or at the least to dismiss it as irrelevant to properly semantic concerns. In the most extreme cases, the attitude of anti-pragmatist philosophers of language to Wittgenstein's picture verges on that of the Victorian lady to Darwin's theory: one hopes that it is not true, and that if it is true, at least that it not become generally known.

But I do not think we are obliged to choose between these approaches. They should be seen as complementing rather than competing with one another. Semantics and pragmatics, concern with meaning and concern with use, ought surely to be understood as aspects of one, more comprehensive, picture of the discursive. Pragmatist considerations do not oblige us to focus on pragmatics to the exclusion of semantics; we can deepen our semantics by the addition of pragmatics. If we extract consequences from the pragmatists' observations somewhat more modestly and construe the analytic project somewhat more broadly, the two will be seen not only as compatible, but as mutually illuminating. If we approach the pragmatists' observations in an analytic spirit, we can understand pragmatics as providing special resources for extending and expanding the analytic semantic project: extending it from exclusive concern with relations among meanings to encompass also relations between meaning and use. In its most ambitious form, as in the present project, such an enterprise would aspire to articulate something like a *logic* of the relations between meaning and use.

If we leave open the possibility that the use of some vocabulary may be illuminated by taking it to express some sort of meaning or content—that is, if we do not from the beginning embrace theoretical semantic nihilism—then the most important positive pragmatist insight will be one complementary to the methodological pragmatism I have already identified. The thought underlying the pragmatist line of thought is that what makes some bit of vocabulary mean what it does is how it is used. What we could call *semantic* pragmatism is the view that the only explanation there could be for how a given *meaning* gets associated with a vocabulary is to be found in the *use* of that vocabulary: the practices by which that meaning is conferred or the abilities whose exercise constitutes deploying a vocabulary with that meaning. To broaden the classical project of analysis in the light of the pragmatists' insistence on the centrality of pragmatics, we can focus on this fundamental relation between use and meaning, between practices or practical abilities and vocabularies. We must look at what it is to use locutions *as* expressing meanings—that is, at what one must *do* in order to count as *saying* what the vocabulary lets practitioners express. I am going to call this kind of relation "practice-vocabulary sufficiency"—or, usually, "PV-sufficiency" for short. It obtains when engaging in a specified set of practices or exercising a specified set of abilities[8] is sufficient for someone to count as *deploying* a specified vocabulary.

Of course it matters a lot how we think about these content-conferring, vocabulary-deploying practices or abilities. The semantic pragmatist's claim that use confers meaning (so talk of practices or the exercise of abilities as deploying vocabularies) reverts to triviality if we are allowed to talk about "using the tilde to express <u>negation</u>," "the ability to mean <u>red</u> by the word 'red'," or "the capacity to refer to electrons by the word 'electron'," (or, I think, even *intentions* so to refer). And that is to say that the interest of the PV-sufficiency of some set of practices or abilities for the deploying of a vocabulary is quite sensitive to the *vocabulary* in which we *specify* those practices-or-abilities. Talk of practices-or-abilities has a definite sense only insofar as it is relativized to the vocabulary in which those practices-or-abilities are specified. And that means that besides

---

[8] For the purposes of the present project, I will maintain a studied neutrality between these options. The apparatus I am introducing can be non-committal as to whether we understand content-conferring *uses* of expressions in terms of social practices or individual abilities, or some more complicated constellation of both.

PV-sufficiency, we should consider a second basic meaning-use relation: "vocabulary-practice sufficiency," or just "*VP-sufficiency*," is the relation that holds between a vocabulary and a set of practices-or-abilities when that vocabulary is sufficient to *specify* those practices-or-abilities. VP-sufficient vocabularies that *specify* PV-sufficient practices let one *say* what it is one must *do* to count as engaging in those practices or exercising those abilities, and so to deploy a vocabulary to *say* something.

PV-sufficiency and VP-sufficiency are two basic *meaning-use relations* (MURs). In terms of those basic relations, we can define a more complex relation: the relation that holds between vocabulary V′ and vocabulary V when V′ is VP-sufficient to specify practices-or-abilities P that are PV-sufficient to deploy vocabulary V. This VV-relation is the *composition* of the two basic MURs. When it obtains I will say that V′ is a *pragmatic metavocabulary* for V. It allows one to *say* what one must *do* in order to count as *saying* the things expressed by vocabulary V. We can present this relation graphically in a *meaning-use diagram* (MUD), as shown in Figure 1.1.

The conventions of this diagram are:

- Vocabularies are shown as ovals, practices-or-abilities as (rounded) rectangles.
- Basic meaning-use relations are indicated by solid arrows, numbered and labeled as to kind of relation.
- Resultant meaning-use relations are indicated by dotted arrows, numbered and labeled as to kind and the basic MURs from which they result.

The idea is that a resultant MUR is the relation that obtains when all of the basic MURs listed on its label obtain.
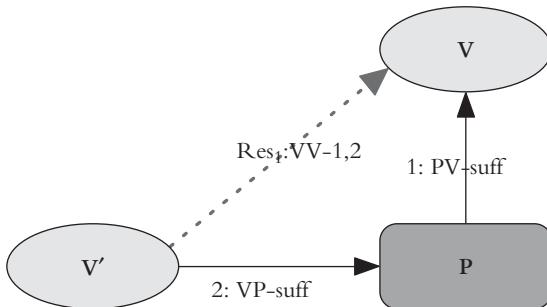

Figure 1.1 Meaning-use diagram 1: pragmatic metavocabulary

Being a pragmatic metavocabulary is the simplest species of the genus I want to introduce here. It is a *pragmatically mediated semantic relation* between vocabularies. It is pragmatically mediated by the practices-or-abilities that are *specified* by one of the vocabularies (which *say* what counts as *doing* that) and that *deploy* or are the *use* of the other vocabulary (what one says *by* doing that). The semantic relation that is established thereby between the two vocabularies is of a distinctive sort, quite different from, for instance, definability, translatability, reducibility, and supervenience. **My basic suggestion for extending the classical project of *analysis* so as to incorporate as essential positive elements the insights that animate the *pragmatist* critique of that project is that, alongside the classical semantic relations between vocabularies that project has traditionally appealed to, we consider also *pragmatically mediated* ones—of which the relation of being a pragmatic metavocabulary is a paradigm**. I will introduce an apparatus that recursively generates an infinite set of such pragmatically mediated semantic relations. In fact, I will eventually argue that unless we take steps along these lines, we cannot properly understand the expressive roles played by some of the kinds of vocabulary with which the analytic tradition has been most centrally concerned: *logical*, *modal*, *normative*, and *intentional* vocabularies.

Under what circumstances would this simplest pragmatically mediated semantic relation be philosophically interesting, when considered in connection with the sorts of vocabularies that have been of most interest to classical analysis? At least one sort of result that could be of considerable potential significance, I think, is if it turned out that, in some cases, pragmatic metavocabularies exist that differ significantly in their expressive power from the vocabularies for the deployment of which they specify sufficient practices-or-abilities. I will call that phenomenon "*pragmatic expressive bootstrapping*." If one vocabulary is strictly weaker in expressive power than the other, I will call that *strict* expressive bootstrapping. We are familiar with this sort of phenomenon in ordinary semantics, where sometimes a semantic metalanguage differs substantially in expressive power from its object language—for instance, where we can produce an extensional metalanguage for intensional languages, as in the case of possible worlds semantics for modality. One example of a claim of this shape in the case of pragmatically mediated semantic relations—though of

course it is not expressed in terms of the machinery I have been introducing—is Huw Price's pragmatic normative naturalism.[9] He argues, in effect, that although normative vocabulary is not *reducible to* naturalistic vocabulary, it might still be possible to *say* in wholly naturalistic vocabulary what one must *do* in order to be *using* normative vocabulary. If such a claim about the existence of an expressively bootstrapping naturalistic pragmatic metavocabulary for normative vocabulary could be made out, it would evidently be an important chapter in the development of the naturalist core program of the classical project of philosophical analysis. It would be a paradigm of the sort of payoff we could expect from extending that analytic project by including pragmatically mediated semantic relations.

The meaning-use diagram of the pragmatically mediated semantic relation of being a pragmatic metavocabulary illustrates a distinctive kind of *analysis* of that relation. It exhibits that relation as the resultant, by composition, of the two basic meaning-use relations of PV-sufficiency and VP-sufficiency. A complex MUR is analyzed as the product of operations applied to basic MURs. This is *meaning-use analysis*. The same analytic apparatus applies also to more complex pragmatically mediated semantic relations. Consider one of the pragmatist criticisms that Sellars addresses to the empiricist core program of the classical analytic project. It turns on the assertion of the *pragmatic dependence* of one set of vocabulary-deploying practices-or-abilities on another.

Because he thinks part of what one is *doing* in saying how things merely appear is withholding a commitment to their actually being that way, and because one cannot be understood as *withholding* a commitment that one cannot *undertake*, Sellars concludes that one cannot have the ability to say or think how things *seem* or *appear* unless one also has the ability to make claims about how things *actually are*. In effect, this Sellarsian pragmatist critique of the phenomenalist form of empiricism consists in the claim that the practices that are PV-sufficient for 'is'-ϕ talk are PP-necessary for the practices that are PV-sufficient for 'looks'-ϕ talk.[10] That pragmatic dependence of practices-or-abilities then induces a resultant pragmatically

---

[9] See his ''Naturalism without Representationalism,'' in Mario de Caro and David Macarthur (eds.), *Naturalism in Question* (Harvard University Press, 2004), 71–90.

[10] I discuss this argument in greater detail in the final chapter of *Tales of the Mighty Dead* (Harvard University Press, 2002).
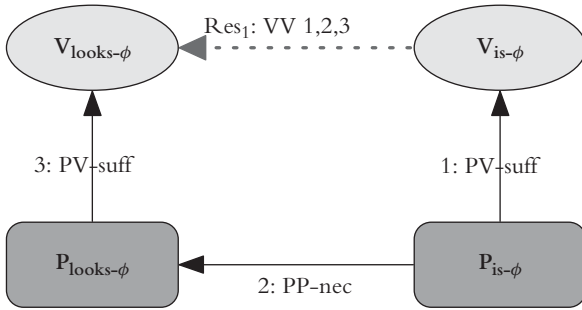
Figure 1.2 Meaning-use diagram 2: pragmatically mediated semantic presupposition

mediated semantic relation between the vocabularies. The meaning-use diagram for this claim is shown in Figure 1.2.

The resultant MUR here is a kind of complex, pragmatically mediated VV-necessity, or semantic presupposition.

In fact, although Sellars's argument for the crucial PP-necessity relation of pragmatic dependence of one set of vocabulary-deploying practices-or-abilities on another is different, his argument against the observational version of empiricism—the claim that purely non-inferential, observational uses do not form an autonomous discursive practice, but presuppose inferential uses—has exactly the same form (Figure 1.3).

For these cases, we can say something further about the nature of the pragmatically mediated semantic relation that is analyzed as the resultant MUR in these diagrams. For, instead of jumping directly to this VV resultant MUR, we could have put in the composition of the PP-necessity
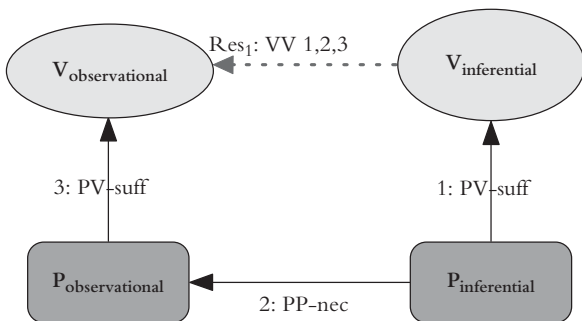
Figure 1.3 Meaning-use diagram 3: pragmatically mediated semantic presupposition
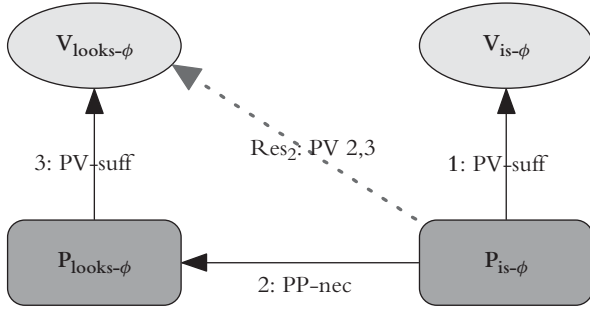
Figure 1.4 Meaning-use diagram 4: composition

and second PV-sufficiency relation, yielding a kind of complex pragmatic presupposition (Figure 1.4).

If this diagram were completed by an arrow from $V_{is-\phi}$ to $V_{looks-\phi}$ such that the *same* diagonal resultant arrow could represent *both* the composition of relations 2 and 3 *and* the composition of relation 1 and the newly supplied one, then category theorists would say that the diagram *commutes*. And the arrow that needs to be supplied to make the diagram commute they call the *retraction* of relation 1 through the composition Res_2 (Figure 1.5).

After composition, then, a slightly more complex form of resultant MUR is retraction. Analyzing the structure of Sellars's pragmatist arguments against empiricism requires recognizing the pragmatically mediated semantic relation he claims holds between phenomenal and objective vocabulary as the retraction of a constellation of more basic meaning-use relations.
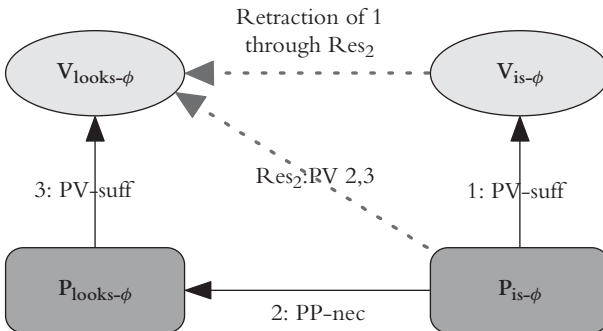
Figure 1.5 Meaning-use diagram 5: composition and retraction

4 Automata: Syntactic PV-sufficiency and VP-sufficiency

Now this is all extremely abstract. To make it more definite, we need to fill in (at least) the notions of <u>vocabulary</u>, <u>practice-or-ability</u>, <u>PV-sufficiency</u>, and <u>VP-sufficiency</u>, which are the fundamental elements that articulate what I am calling the "meaning-use analysis" of resultant meaning-use relations—in particular, the pragmatically mediated semantic relations between vocabularies that I am claiming we must acknowledge in order to pursue the classical project of philosophical analysis in the light of what is right about the pragmatist critique of it. We can begin to do that by looking at a special case in which it is possible to be unusually clear and precise about the things and relations that play these metatheoretic roles. This is the case where 'vocabulary' takes a purely *syntactic* sense. Of course, the cases we eventually care about—and will be discussing in the remaining lectures—involve vocabularies understood in a sense that includes their *semantic* significance. But besides the advantages of clarity and simplicity, we will find that some important lessons carry over from the syntactic to the semantic case.

The restriction to vocabularies understood in a spare syntactic sense leads to correspondingly restricted notions of what it is to *deploy* such a vocabulary, and what it is to *specify* practices-or-abilities sufficient to deploy one. Suppose we are given an *alphabet*, which is a finite set of primitive sign types—for instance, the letters of the English alphabet. The *universe* generated by that alphabet then consists of all the finite strings that can be formed by concatenating elements drawn from the alphabet. A *vocabulary* over such an alphabet—in the syntactic sense I am now after—is then any subset of the universe of strings that alphabet generates. If the generating alphabet is the English alphabet, then the vocabulary might consist of all English sentences, all possible English texts, or all and only the sentences of *Making It Explicit*.[11]

---

[11] Computational linguists, who worry about vocabularies in this sense, have developed meta-languages for specifying important classes of such vocabularies: the syntactic analogues of semantic metalanguages in the cases we will eventually address. So, for instance, for the alphabet {a,b}, '$a^n b^n$' characterizes the vocabulary that comprises all strings of some finite number of 'a's followed by the *same* number of 'b's. 'a(ba)*b' characterizes the vocabulary that comprises all strings beginning with an 'a', ending with a 'b', and having *any* number of repetitions of the sub-string 'ba' in between.

What can we say about the abilities that count as *deploying* a vocabulary in this spare syntactic sense?[12] The abilities in question are the capacity to *read* and *write* the vocabulary. In this purely syntactic sense, 'reading' it means being able practically to *distinguish* within the universe generated by the alphabet, strings that do, from those that do not, belong to the specified vocabulary. And 'writing' it means practically being able to *produce* all and only the strings in the alphabetic universe that do belong to the vocabulary.

We assume as primitive abilities the capacities to read and write, in this sense, the alphabet from whose universe the vocabulary is drawn—that is, the capacity to respond differentially to alphabetic tokens according to their type, and to produce tokens of antecedently specified alphabetic types. Then the abilities that are PV-sufficient to deploy some vocabularies can be specified in a particularly simple form. They are *finite-state automata* (FSAs). As an example, suppose we begin with the alphabet {a, h, o, !}. Then we can consider the *laughing Santa vocabulary*, which consists of strings such as 'hahaha!', 'hohoho!', 'hahahoho!', 'hohoha!', and so on.[13] Figure 1.6 is a graphical representation of a *laughing Santa finite-state automaton*, which can read and write the laughing Santa vocabulary. The numbered nodes represent the *states* of the automaton, and the alphabetically labeled arcs represent *state-transitions*. By convention, the starting state is represented by a square (State 1), and the final state by a circle with a thick border (State 4).

As a *reader* of the laughing Santa vocabulary, the task of this automaton is to process a finite string, and determine whether or not it is a licit string of the vocabulary. It processes the string one alphabetic character at a time, beginning in State 1. It recognizes the string if and only if (when and only when) it arrives at its final state, State 4. If the first character of the string is not an 'h', it remains stuck in State 1, and rejects the string. If the first
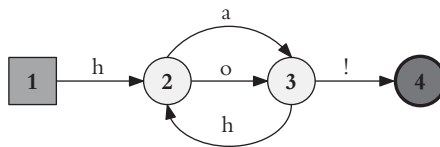
Figure 1.6 The laughing Santa automaton

---

[12] Here we can safely just talk about *abilities*, without danger of restricting the generality of the analysis.

[13] In the syntactic metalanguage for specifying vocabularies that I mentioned in the note above, this is the vocabulary (ha/ho)*!.

character is an 'h', it moves to State 2, and processes the next character. If that character is not an 'a' or an 'o', it remains stuck in State 2, and rejects the string. If the character is an 'a' or an 'o', it moves to State 3. If the next character is an exclamation point, it moves to State 4, and recognizes the string 'ha!' or 'ho!'—the shortest ones in the laughing Santa vocabulary. If, instead, the next character is an 'h', it goes back to State 2, and repeats itself in loops of 'ha's and 'ho's any number of times until an exclamation point is finally reached, or it is fed a discordant character.

As a *writer* of the laughing Santa vocabulary, the task of the automaton is to produce only licit strings of that vocabulary, by a process that can produce any and all such strings. It begins in its initial state, State 1, and emits an 'h' (its only available move), changing to State 2. In this state, it can produce either an 'a' or an 'o'—it selects one at random[14]—and goes into State 3. In this state, it can either tack on an exclamation point, and move into its final state, State 4, finishing the process, or emit another 'h' and return to State 2 to repeat the process. In any case, whenever it reaches State 4 and halts, the string it has constructed will be a member of the laughing Santa vocabulary.

I hope this brief rehearsal makes it clear how the constellation of nodes and arrows that makes up this directed graph represents the abilities to read and write (recognize and produce arbitrary strings of) the laughing Santa vocabulary.[15] What it represents is abilities that are *PV-sufficient* to

---

[14] In fact, it can be shown that every vocabulary readable/writeable by a non-deterministic FSA—such as the laughing Santa automaton—is also readable/writeable by a deterministic one.

[15] For practice, or to test one's grip on the digraph specification of FSAs, consider what vocabulary over the same alphabet that produces the laughing Santa is recognized/produced by this automaton (the "I'll have what she's having" automaton):
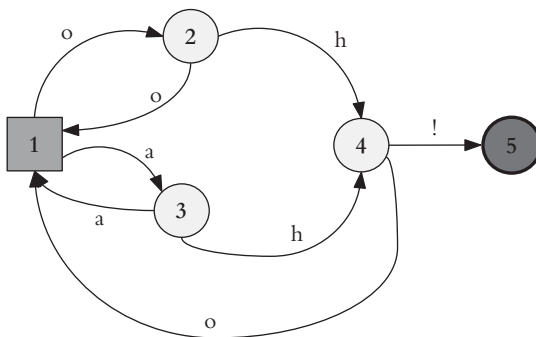
deploy that vocabulary—that is, read and write it, in the attenuated sense appropriate to this purely syntactic case. And the digraph representation is itself a *vocabulary* that is *VP-sufficient* to *specify* those vocabulary-deploying abilities. That is, the digraph representation of this finite-state automaton is a *pragmatic metavocabulary* for the laughing Santa vocabulary. The relation between the digraph vocabulary and the laughing Santa vocabulary is, then, a *pragmatically mediated*—not now *semantic*, but *syntactic*—relation between vocabularies.

It may seem that I am stretching things by calling the digraph form of representation a 'vocabulary'. It will be useful, as a way of introducing my final point in the vicinity, to consider a different form of pragmatic metavocabulary for the laughing Santa vocabulary. Besides the digraph representation of a finite-state automaton, we can also use a *state-table* representation. The state-table for the laughing Santa automaton (LSA) is shown in Table 1.1.

In read mode, the automaton starts in State 1. To see what it will do if fed a particular character, we look at the row labeled with that character. The LSA will Halt if the input string starts with anything other than an 'h', in which case it will change to State 2. In that state, the automaton specified by the table will halt unless the next character is an 'a' or an 'o', in which case it changes to State 3, and so on. (There is no column for State 4, since it is the final state, and accepts/produces no further characters.) Clearly there is a tabular representation corresponding to any digraph representation of an FSA, and vice versa. Notice, further, that we need not use a two-dimensional table to convey this information. We could put the rows one after another, in the form:

aHalt3Halth2Halt2oHalt3Halt!HaltHalt4

Table 1.1 State-table for the laughing Santa automaton

Laughing Santa	State 1	State 2	State 3
a	Halt	3	Halt
h	2	Halt	2
o	Halt	3	Halt
!	Halt	Halt	4
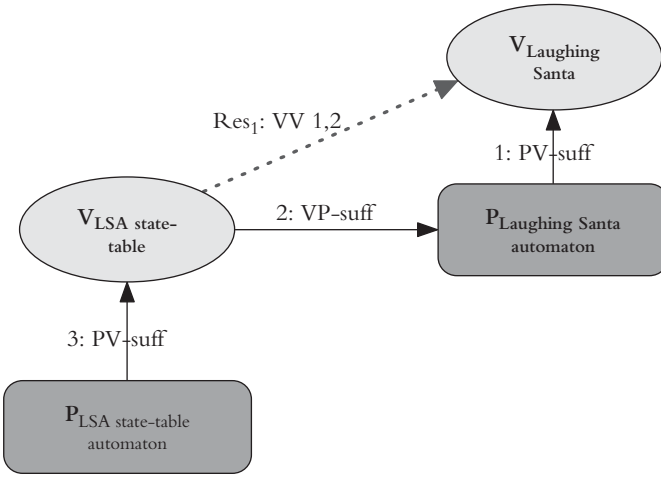
Figure 1.7 Meaning-use diagram 6: specifying the automaton that deploys the laughing Santa vocabulary

This is just a string, drawn from a universe generated by the alphabet of the LSA, together with 'Halt' and the designations of the states of that automaton. The strings that specify FSAs that deploy vocabularies defined over the same basic alphabet as the LSA then form a vocabulary in the technical syntactic sense we have been considering. And that means we can ask about the automata that can read and write *those* state-table encoding vocabularies. The meaning-use diagram for this situation is then as shown in Figure 1.7.

5 The Chomsky hierarchy and a syntactic example of pragmatic expressive bootstrapping

Restricting ourselves to a purely syntactic notion of a <u>vocabulary</u> yields a clear sense of 'pragmatic metavocabulary': both the digraph and the state-table vocabularies are VP-sufficient to specify practical abilities articulated as a finite-state automaton that is PV-sufficient to deploy—in the sense of recognizing and producing—the laughing Santa vocabulary, as well as many others. (Of course, it does that only against the background of a set of abilities PV-sufficient to deploy *those* vocabularies.) Perhaps surprisingly, it also offers a prime example of *strict pragmatic expressive bootstrapping*. For

in this setting we can *prove* that one vocabulary that is expressively weaker than another can nonetheless serve as an adequate *pragmatic* metavocabulary for that stronger vocabulary. That is, even though one can*not say* in the weaker vocabulary everything that can be *said* in the stronger one, one can still *say* in the weaker one everything that one needs to be able to *do* in order to deploy the stronger one.

Here the relevant notion of the <u>relative expressive power</u> of vocabularies is also a purely syntactic one. Already in the 1950s, Chomsky offered mathematical characterizations of the different sets of strings of characters that could be generated by different classes of grammars (that is, in my terms, characterized by different kinds of syntactic metavocabularies) and computed by different kinds of automata. The kinds of vocabulary, grammar, and automata line up with one another, and can be arranged in a strict expressive hierarchy: the Chomsky hierarchy. It is summarized in Table 1.2.

The point I want to make fortunately does not require us to delve very deeply into the information summarized in this table. A few basic points will suffice. The first thing to realize is that not all vocabularies in the syntactic sense we have been pursuing can be read and written by FSAs. For instance, it can be shown that no FSA is PV-sufficient to deploy the vocabulary $a^n b^n$, defined over the alphabet {a,b}, which consists of all strings of any arbitrary number of 'a's followed by the same number of 'b's. The idea behind the proof is that in order to tell whether the right number of 'b's follow the 'a's (when reading) or to produce the right number of 'b's (when writing), the automaton must somehow keep track of how many

Table 1.2 The Chomsky hierarchy

Vocabulary	Grammar	Automaton
Regular	$A \rightarrow aB$ $A \rightarrow a$	Finite state automaton
Context-free	$A \rightarrow$ <anything>	Push–down automaton
Context-sensitive	$c_1 A c_2 \rightarrow c_1$ <anything> c_2	Linear bounded automaton
Recursively enumerable	No restrictions on rules	Turing machine (= 2 Stack PDA)

'a's have been processed (read or written). The only way an FSA can store information is by being in one state rather than another. So, it could be in one state—or in one of a class of states—if one 'a' has been processed, another if two have, and so on. But, by definition, an FSA only has a finite number of states, *and that number is fixed in advance* of receiving its input or producing its output. Whatever that number of states is, and whatever system it uses to code numbers into states (it need not be one-to-one—it could use a decimal coding, for instance), there will be some number of 'a's that is so large that the automaton runs out of states before it finishes counting. But the vocabulary in question consists of arbitrarily long strings of 'a's and 'b's. In fact, it is possible to say exactly which vocabularies FSAs (specifiable by digraphs and state-tables of the sort illustrated above) *are* capable of deploying. These are called the 'regular' vocabularies (or languages).

The next point is that slightly more complex automata *are* capable of deploying vocabularies, such as $a^n b^n$, that are not regular, and hence cannot be read or written by FSAs. As our brief discussion indicated, intuitively the problem that FSAs have with languages like $a^n b^n$ is that they lack *memory*. If we give them a memory, we get a new class of machines: (non-deterministic[16]) *push-down automata* (PDAs). In addition to being able to respond differentially to, and produce tokenings of, the alphabetic types, and being able to change state, PDAs can *push* alphabetic values to the top of a *memory-stack*, and *pull* such values from the top of that stack. PDAs can do everything that FSAs can do, but they can also read and write many vocabularies, such as $a^n b^n$, that are *not* regular and so cannot be read and written by FSAs. The vocabularies they can deploy are called "context-free." All regular vocabularies are context-free, but not vice versa. This proper containment of classes of vocabularies provides a clear sense, suitable to this purely syntactic setting, in which one vocabulary can be thought of as "expressively more powerful" than another: the different kinds of grammar can specify, and the different kinds of automata can compute, ever larger classes of vocabularies. Context-free vocabularies that are not regular require more powerful grammars to specify them, as well as more powerful automata to deploy them. FSAs are special kinds of PDAs, and all

---

[16] By contrast to FSAs, there need not in general be for every vocabulary computable by a non-deterministic PDA, some deterministic PDA that reads and writes the same vocabulary.

the automata are special kinds of Turing machines. Recursively enumerable vocabularies are not, in general, syntactically reducible to context-sensitive, context-free, or regular ones. And the less capable automata cannot read and write all the vocabularies that can be read and written by Turing machines.

Nonetheless, if we look at *pragmatically mediated* relations between these syntactically characterized vocabularies, we find that they make possible a kind of *strict expressive bootstrapping* that permits us in a certain sense to evade the restrictions on expressive power enforced for purely syntactic relations between vocabularies. The hierarchy dictates that only the abilities codified in Turing machines—two-stack push-down automata—are *PV-sufficient* to *deploy* recursively enumerable vocabularies in general. But now we can ask: what class of languages is *VP-sufficient* to *specify* Turing machines, and hence to serve as sufficient *pragmatic* metavocabularies for recursively enumerable vocabularies in general? The surprising fact is that **the abilities codified in Turing machines—the abilities to recognize and produce arbitrary recursively enumerable vocabularies—can quite generally be specified in *context-free* vocabularies**. It is demonstrable that context-free vocabularies are strictly weaker in syntactic expressive resources than recursively enumerable vocabularies. The PDAs that can read and write only context-free vocabularies cannot read and write recursively enumerable vocabularies in general. But it is possible to *say* in a context-free vocabulary what one needs to be able to *do* in order to deploy recursively enumerable vocabularies in general.

The proof of this claim is tedious but not difficult, and the claim itself is not at all controversial—though computational linguists make nothing of it, having theoretical concerns very different from those that lead me to underline this fact. (My introductory textbook leaves the proof as an exercise to the reader.[17]) General-purpose computer languages such as Pascal and C++ can specify the algorithms a Turing machine, or any other universal computer, uses to compute any recursively enumerable function, hence to recognize or produce any recursively enumerable vocabulary. And they are invariably context-free languages[18]—in no small part just

---

[17] Thomas A. Sudkamp, *Languages and Machines*, 2nd ed. (Addison Wesley Longman, 1998), chapter 10.

[18] In principle. There are subtleties that arise when we look at the details of actual implementations of particular computer languages, which can keep them from qualifying as strictly context-free.
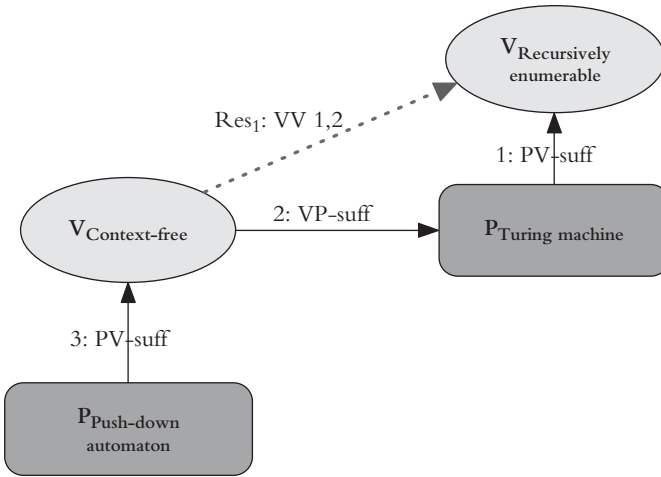
Figure 1.8 Meaning-use diagram 7: syntactic pragmatic expressive bootstrapping

because the simplicity of this type of grammar makes it easy to write parsers for them. Yet they suffice to specify the state-table, contents of the tape (or of the dual stacks), and primitive operations of any and every Turing machine. Figure 1.8 shows the MUD characterizing this pragmatically mediated relation between syntactically characterized vocabularies.

I called the fact that context-free vocabularies can be adequate pragmatic metavocabularies for recursively enumerable vocabularies in general 'surprising', because of the provable syntactic irreducibility of the one class of vocabularies to the other. But if we step back from the context provided by the Chomsky hierarchy, we can see why the possibility of such pragmatic expressive bootstrapping should not, in the end, be surprising: for all the result really means is that context-free vocabularies let one *say* what it is one must *do* in order to say things they cannot themselves say, because the ability to deploy those context-free vocabularies does not include the abilities those vocabularies let one specify. Thus, for instance, there is no reason that an FSA could not read and write a vocabulary that included commands such as ''Push an 'a' onto the stack''—and thus specify the program of a PDA—even though it itself has no stack, and could not *do* what the vocabulary it is deploying specifies. A coach might be able to tell an athlete exactly what to do, and even how to do it, even though the coach cannot himself do what he is telling the athlete to do, does not have the abilities he is specifying. We ought not to boggle at the possibility

of an expressively weaker pragmatic metavocabulary having the capacity to say what one must do in order to deploy an expressively stronger one. We should just look to see where this seems in fact to be possible for vocabularies we care about, and what we can learn from such relations when they do obtain.

6 Looking ahead

Let us recall what motivated this rehearsal of some elements of automaton theory and introductory computational linguistics. I suggested that a way to extend the classical project of semantic analysis so as to take account of the insights of its pragmatist critics is to look analytically at relations between meaning and use. More specifically, I suggested focusing to begin with on two in some sense complementary relations: the one that holds when some set of practices-or-abilities is PV-sufficient to *deploy* a given vocabulary, and the one that holds when some vocabulary is VP-sufficient to *specify* a given set of practices-or-abilities. The composition of these is the simplest *pragmatically mediated semantic relation* between vocabularies: the relation that holds when one vocabulary is a sufficient pragmatic metavocabulary for another. It is a paradigm of the infinite, recursively generable class of complex, pragmatically mediated semantic relations that I propose to lay alongside the other semantic relations between vocabularies that have been investigated by analytic philosophers (for instance, those who address the core programs of empiricism and naturalism): relations such as analyzability, definition, translation, reduction, truth-making, and supervenience. I suggested further that pragmatic metavocabularies might be of particular interest in case they exhibited what I called "expressive bootstrapping"—cases, that is, in which the expressive power of the pragmatic metavocabulary differs markedly from that of the target vocabulary, most strikingly when the metavocabulary is substantially expressively weaker—a phenomenon Tarski has led us *not* to expect for semantic metavocabularies.

We have now seen that all of these notions can be illustrated with particular clarity for the special case of purely syntactically characterized vocabularies. The abilities that are PV-sufficient to deploy those vocabularies, in the sense of the capacity to recognize and produce them, can be thought of as various sorts of automata. There are several well-established,

different-but-equivalent vocabularies that are known to be VP-sufficient to specify those automata. In this special syntactic case we can accordingly investigate the properties of pragmatic metavocabularies, and when we do, we find a striking instance of *strict expressive bootstrapping* in a pragmatically mediated syntactic relation between vocabularies.

Of course, the cases we really care about involve *semantically* significant vocabularies. Are there any interesting instances of these phenomena in such cases? I have indicated briefly how some of Sellars's pragmatist criticisms of various ways of pursuing the empiricist program can be understood to turn on pragmatically mediated semantic relations. And I mentioned Huw Price's idea that although normative vocabulary is not semantically reducible to naturalistic vocabulary, naturalistic vocabulary might suffice to specify what one must *do*—the practices-or-abilities one must engage in or exercise—in order to deploy normative vocabulary. Here is another example that I want to point to, though I cannot develop the claim here. For roughly the first three-quarters of the twentieth century, philosophers who thought about indexical vocabulary took for granted some version of the doctrine that a tokening *n* of an expression of the type 'now' was *synonymous* with, definable or semantically analyzable as, 'the time of utterance of *n*', and similarly for 'here' and 'the place of utterance of *h*', and so on. During the 1970s, philosophers such as John Perry, David Lewis, and G. E. M. Anscombe, by focusing on the use of indexicals in modal and epistemic contexts, showed decisively that this cannot be right: what is expressed by indexical vocabulary cannot be expressed equivalently by non-indexical vocabulary. This fact seems so obvious to us now that we might be led to wonder what philosophers such as Russell, Carnap, and Reichenbach could have been thinking for all those years. I want to suggest that the genuine phenomenon in the vicinity is a *pragmatically mediated* semantic relation between these vocabularies. Specifically, in spite of the *semantic* irreducibility of indexical to non-indexical vocabulary, it is possible to *say*, entirely in non-indexical terms, what one must *do* in order to be deploying indexical vocabulary correctly: to be saying essentially and irreducibly indexical things. For we can formulate practical rules such as:

- If, at time *t* and place $<x, y, z>$, speaker *s* wants to assert that some property *P* holds of $<x, y, z, t, s>$, it is correct to say "P holds of *me, here* and *now*." And

- If a speaker s at time t and place $<x, y, z>$ asserts "P holds of *me, here* and *now*," the speaker is committed to the property P holding of $<x, y, z, t, s>$.

Further (as I show in the appendix to the next lecture, where the necessary concepts have been introduced), those responses can be algorithmically elaborated so as to play the role distinctive of *essential* indexicals. Non-indexical vocabulary can serve as an adequate *pragmatic* metavocabulary for indexical vocabulary. The fact that one nonetheless cannot *say* in non-indexical terms everything that one can *say* with indexical vocabulary just shows that these vocabularies have different expressive powers, so that the pragmatically mediated semantic relation between them is a case of strict pragmatic expressive bootstrapping.

In the lectures to come, I will be doing three things:

- further developing the conceptual apparatus of meaning-use analysis, by introducing both new basic meaning-use relations and new combinations of them;
- applying that apparatus to vocabularies of ongoing philosophical interest (logical, modal, normative, intentional); and
- seeing what new pragmatically mediated semantic relations become visible in that way.

Each subsequent lecture will report some further unexpected, suggestive results, which fit together cumulatively to constitute a distinctive, novel picture of what we would previously have thought was familiar terrain.

Besides pragmatically mediated semantic relations between vocabularies, there is another sort of pragmatic analysis, which relates one constellation of practices-or-abilities to another. It corresponds to another basic meaning-use relation: the kind of PP-sufficiency that holds when having acquired one set of abilities means one can already do everything one needs to do, *in principle*, to be able to do something else. One concrete way of filling in a definite sense of "in principle" is by *algorithmic elaboration*, where exercising the target ability just is exercising the right basic abilities in the right order and under the right circumstances. As an example, the ability to do long division just consists in exercising the abilities to do multiplication and subtraction and write down the results of those calculations, according to a particular conditional branched-schedule algorithm. The practical abilities that *implement* such an algorithmic PP-sufficiency relation are just

those exercised by a finite-state *automaton*. Indeed, automata are defined by a definite set of *meta-abilities*: abilities to *elaborate* a set of primitive abilities into a set of more complex ones, which can accordingly be pragmatically *analyzed* in terms of or *decomposed* into the other.

To get a usefully general concept of the PP-sufficiency of a set of basic abilities for a set of more complex ones, we need to move beyond the purely *syntactic* automata I have described so far. One way to do that is to replace their specialized capacities to read and write symbols—in the minimal sense of classifying tokens as to types and producing tokens of specified types—by more general recognitional and productive capacities. These are abilities to respond differentially to various non-symbolic stimuli (for instance, the visible presence of red things), corresponding to reading, and to respond by producing performances of various non-symbolic kinds (for instance, walking north for a mile), corresponding to writing. What practically implements the algorithmic elaboration of such a set of basic differential responsive abilities is a finite state *transducing* automaton.

In my third lecture, I will argue that the notion of the <u>algorithmic decomposability</u> of some practices-or-abilities into others that results suggests in turn a pragmatic generalization of the classical program of artificial intelligence (AI) functionalism—which, though a latecomer in the twentieth century, deserves, I think, to count as a *third* core program of classical semantic analysis. AI functionalism traditionally held itself hostage to a commitment to the purely *symbolic* character of intelligence in the sense of sapience. But broadening our concern from automata as purely syntactic engines to the realm of *transducing* automata puts us in a position to see AI functionalism as properly concerned with the algorithmic decomposability of discursive (that is, vocabulary-deploying) practices-and-abilities. What I will call the "pragmatic" thesis of artificial intelligence is the claim that the ability to engage in some autonomous discursive practice—a language game one could play though one played no other—can be *algorithmically decomposed* into non-discursive abilities. The arguments for and against this pragmatic version of AI-functionalism look quite different from those arrayed on the opposing sides of the debate about the prospects of symbolic AI.

The notion of <u>PP-sufficiency</u> brings into view a slightly more complicated pragmatically mediated semantic relation between vocabularies: that which obtains when practices PV-sufficient for V_1 are PP-sufficient

(in the sense that they can be algorithmically elaborated into) practices PV-sufficient for V_2. Another basic meaning-use relation of the kind we have been considering is PV-*neccessity*, the converse of PV-sufficiency. It obtains when one cannot deploy a certain vocabulary without engaging in the specified practice, or exercising the specified ability. For example, I have argued elsewhere that nothing could count as engaging in an autonomous discursive practice (hence using a vocabulary one could use though one used no other) that did not include *asserting* and *inferring*. Considering that basic MUR permits the formulation of a complex resultant MUR that is a variant on the prior one: a relation that obtains where practices PV-*necessary* for V_1 are PP-sufficient for practices-or-abilities PV-sufficient for V_2.

It can happen, I will argue, that such a V_2 is also VP-sufficient to *specify* the practices-or-abilities that are PV-sufficient to deploy V_1. A MUD for this is shown in Figure 1.9.

In my next lecture, I will introduce a version of this complex resultant pragmatically mediated semantic relation (what I call, for short, being "universally LX"), and argue that it constitutes the genus of which *logical* vocabulary is a species. More specifically, I will argue that logical vocabulary both can be algorithmically elaborated from and is explicative of practices that are PV-necessary for the autonomous deployment of any vocabulary at all. And I will argue that the most illuminating way to explain and justify the distinctive privileged role accorded to logical vocabulary by the classical project of philosophical analysis—what I have here called "semantic logicism"—is by appeal to this whole constellation of basic
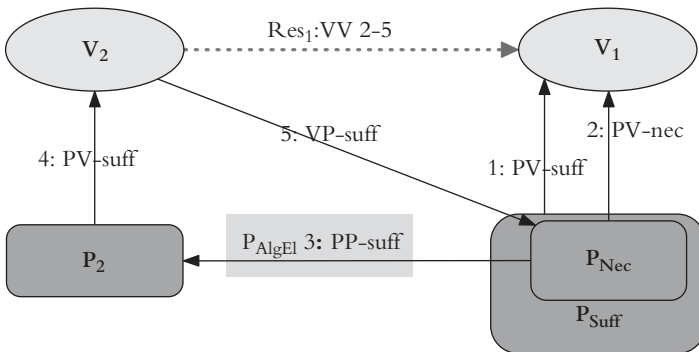

Figure 1.9 LX: V_2 is elaborated from and explicative of practices PV-necessary for V_1

meaning-use relations, and the complex pragmatically mediated semantic relation that results from it.

My last three lectures will address modal vocabulary, normative vocabulary, and the pragmatically mediated semantic relations they stand in to ordinary objective, empirical, and naturalistic vocabularies, to each other, and to what is expressed by intentional vocabulary. The modal revolution in the last third of the twentieth century breathed new life into semantic logicism, providing powerful new expressive tools, which have been of great use to those pursuing naturalistic programs, for instance. But this successor version raises the same question of vindication that I consider for semantic logicism in my second lecture: what justifies according modal concepts this special, privileged role in our semantic analytic enterprise? This question is particularly urgent since the empiricist program had always been—traditionally with Hume, and in the twentieth-century logical form, with Quine—particularly and specifically hostile to and critical of this vocabulary.

I will begin my treatment of modality, in my fourth lecture, with a consideration of this question, and with a vindication of the role of modal vocabulary that parallels the one I will already have offered for ordinary logical vocabulary: modal vocabulary, too, can be elaborated from and is explicative of features integral to every autonomous discursive practice—features intimately related to, but distinct from, those made explicit by ordinary logical vocabulary. I will then enter into an extended treatment of the relation between *alethic* and *deontic* (modal and normative) vocabularies. When we look at those vocabularies through the lens of meaning-use analysis, a sequence of startling relations between them emerges.

For a start, I argue that deontic normative vocabulary is *also* universally LX (that it is VP-sufficient to specify practices-or-abilities that are both PV-necessary for deploying any autonomous vocabulary, and PP-sufficient for practices-or-abilities PV-sufficient for deploying the deontic normative vocabulary that explicates them). Although in this regard it belongs in a box with alethic modal vocabulary, the features of autonomous vocabulary use that it explicates are quite different from those explicated by modal vocabulary. I then argue that what lies behind Sellars's dark and pregnant claim that "the language of modality is a transposed language of norms" is the fact that deontic normative vocabulary can serve as a *pragmatic*

metavocabulary for alethic modal vocabulary. In my fifth lecture, I will show how exploiting that relation makes possible a new kind of *directly modal* formal semantics that makes no appeal to *truth*: incompatibility semantics. It in turn gives us a new semantic perspective both on traditional logical vocabulary, and on modal vocabulary. The final lecture will then weave all these strands into a meaning-use analysis of intentionality itself (what is expressed by intentional vocabulary) as a pragmatically mediated semantic relation essentially involving both what is expressed by modal and what is expressed by normative vocabulary.

The substantive cumulative result of this sequence of revelations about modal and normative vocabulary is to put new flesh on the bones of ideas that originate with Kant, and are developed by his tradition up through the traditional American pragmatists, and are reinterpreted by Sellars in the middle years of the twentieth century. And the methodological result of this development and application of meaning-use analysis is a new synthesis of pragmatism and analytic philosophy—one that shows how concerns and considerations at the heart of the pragmatist critique of semantic analysis can be seen to have been implicitly at work within the analytic tradition all along.

The title of this book, *Between Saying and Doing*, evidently refers to my aspiration to present a new way of thinking about the relations between meaning and use that arises when we think systematically about *saying* what we are *doing* when we are *saying* something. But the phrase itself is taken from an Italian proverb: "Between saying and doing, many a pair of shoes is worn out." Following the argumentative and constructive path I am proposing for exploring the intricate and revealing ways in which semantics and pragmatics interdigitate will require wearing out a few.

2

Elaborating Abilities: The Expressive Role of Logic

1 Automata as algorithmically elaborating abilities

In the previous lecture I began to set out my enterprise against the background of an understanding of the classical project of philosophical analysis as having the task of exhibiting what is expressed semantically by one vocabulary (one sort of meaning) as the logical elaboration of what is expressed by another. From this point of view, empiricism and naturalism show up as generic core analytic programs, with their species distinguished in part by the vocabularies they treat as basic and by those they seek to elaborate on those bases, and in part by the sort of logically articulated elaboration they see as relating the basic and target vocabularies. Pragmatism can be thought of as challenging these analytic undertakings, by insisting that appeal to *meaning* is just *one*, optional, theoretically laden way of dealing with *use*, one that is based on a defective, scientistic, uniformitarian theory of what is required to understand what one has to *do* in order to count as *saying* various kinds of thing—as deploying various vocabularies. But if we accept the methodological pragmatism that tells us that the point of *semantic* theory is to make sense of *pragmatics*—that *meanings* should be thought of as theoretical entities postulated to explain, or at least to codify, proprieties of *use*—and supplement it with the *semantic* pragmatism that tells us that only its use can explain the association of meaning with, or its expression by, a vocabulary, we see that we are not forced to choose between thinking in terms of the *meanings* expressed by vocabularies and thinking in terms of their *use*. I want to show how pragmatism can be turned from a pessimistic, even nihilistic, counsel of theoretical despair into a definite, substantive,

progressive, and promising program in the philosophy of language: indeed, how it can be understood as ushering in a new phase of the analytic project.

A central task must be to get clearer about the notion of <u>use</u> that pragmatism seeks (irenically) to lay alongside semantics, or with which it seeks (martially) to displace semantics. The starting place I suggested last time is the way in which automaton theory lets us specify the abilities needed to *deploy* syntactically characterized vocabularies, in the sense of being able both to distinguish and to generate them. This idiom illustrated a number of basic meaning-use relations: relations between (as I suggested we label them) vocabularies and practices-or-abilities. My discussion of them began with vocabularies that are *syntactically* specified, rather than *semantically* specified—a restriction that pays off in the clarity and definiteness of the relations involved, but only at the high price of abstracting from the dimension of *semantic* expressiveness that makes us care philosophically about vocabularies in the first place. (After all, it is *meaning*-use relations we are ultimately after.) Nonetheless, we can see already at this point that the metavocabulary we use to characterize a vocabulary makes a significant difference to what sort of practices-or-abilities we can count as deploying it. We also saw that the practices-or-abilities to deploy those metavocabularies—the ones that are, in my terms, PV-sufficient for them—must be thought of in turn in terms of the vocabularies in which *they* are specified.

I promised that we could build on this initial, simplified syntactic model of basic and resultant meaning-use relations—especially PV-sufficiency, VP-sufficiency, the pragmatically mediated VV-relation that is their composition, and the sort of pragmatic expressive bootstrapping exhibited by such relations even in the syntactic case—to yield insights into corresponding relations between genuine meanings and uses for vocabularies characterized in full-blooded, *semantic* metavocabularies. So far, I have presented automata as merely syntactic engines, which is to say as sets of primitive PV-sufficient abilities to deploy purely syntactically characterized vocabularies. But there is another way to think about the abilities that the automaton-theoretic VP-sufficient vocabulary specifies. Automata put together primitive abilities so that they add up to more complex ones. Automata are the practical embodiments of *algorithms*. And algorithms generally say how some set of primitive abilities can be so exercised as to constitute more complex abilities. For instance, an algorithm implemented

by an automaton tells us how to put together the capacities to multiply and subtract so as to amount to the capacity to do long division.

Thought of this way, automata are defined by a definite set of *meta-abilities*: abilities to *elaborate* a set of primitive abilities into a set of more complex ones. In the metavocabulary for meaning-use relations I am introducing here, they *implement* PP-sufficiency relations: the kind of relation that obtains when the capacity to engage in one sort of practice or to exercise one sort of ability is *in principle sufficient* for the capacity to engage in other practices, or to exercise other abilities. Putting the point the other way around, an automaton-theoretic specification of how a set of primitive practices-or-abilities can be algorithmically elaborated into a more complex set of practices-or-abilities shows how the latter can be *pragmatically analyzed* into the former. Semantic analysis in the most straightforward case—for instance, Russell's analysis of definite descriptions—shows how some *meanings* can be exhibited as complex, in the sense that they can be understood as compounded by definite means out of more basic meanings. The semantic logicism characteristic of the classical project of philosophical analysis in the twentieth century insists that it is *logical* vocabulary that articulates or makes explicit such semantic compounding. We can lay alongside this conception a notion of *pragmatic analysis*. This is the analysis of some *uses* (rather than meanings) as complex, in the sense that they can be understood as compounded by definite means out of simpler uses, that is, in the regimentation I am employing, practices-or-abilities.[1] My claim is that automata can be thought of as one way of implementing such pragmatic analyses. What thinking about automata in this broad sense will do is to teach us that *algorithmic elaboration* of primitive abilities into complex ones plays the same role in *pragmatic* analysis that *logic* does in *semantic* analysis. Algorithmic elaboration is a kind of logic of practical abilities.

My aim is to weave these two sorts of analysis together as essential elements of a more comprehensive kind of meaning-use analysis. In this lecture I begin to explore what happens when we add the PP-sufficiency relation of algorithmic elaboration, implemented by automata, to the

---

[1] Of course, the issue of how some *doings* (whether intentional or not) can be presented as products, compounded in antecedently specifiable ways, out of more primitive doings, is a *very* general one. But we have the advantage that the doings we care most about are sayings: practices-or-abilities that count as deploying vocabularies, as conferring or applying meanings: *discursive* practices-or-abilities. This gives our approach to the general question a more particular focus.

conceptual apparatus of the PV-sufficiency of practices-or-abilities to deploy vocabularies and the VP-sufficiency of vocabularies to specify practices-or-abilities. Doing that is the first step in the move to understanding the practices-or-abilities involved in deploying *semantically* characterized vocabularies.

2 Transducing automata

Automata, in the general sense in which I want to think about them, are constellations of practices-or-abilities that algorithmically elaborate sets of primitive practices-or-abilities into more complex ones. *Transducing* automata[2] generalize the primitive reading-and-writing abilities of finite-state automata to include *discriminating stimuli* of any kind, on the input side, and *differentially responding* in any way, on the output side. That is, instead of an alphabet of character-types, tokenings of which can be indifferently read or written, these automata are defined over a pair <S, R> of *stimulus*-types and *response*-types. The stimulus-types are any circumstances to which the system is able to respond differentially. The response-types are any kinds of performance that can be differentially elicited from the system—that is, which it produces in some circumstances and not others. A *single-state transducing automaton* (SSTA) can be specified by a state-table that is just a set of pairs of stimulus-kinds and the kinds of responses they would elicit. It is just a set of reliable differential responsive dispositions.

The stimulus-response model might seem to impose a formal, narrowly behaviorist straitjacket on what counts as a primitive ability. Behaviorism is lurking in the vicinity, but I think it is important to see that its proximity does *not* arise from this characterization of primitive abilities. For how restrictive the stimulus-response model of such abilities is depends on the VP-sufficient vocabulary that we are allowed to use in specifying the discriminable stimuli and differentially elicitable performances. If *no* restrictions are imposed, then candidate stimuli one might have the ability to discriminate can include such species as poetry that qualifies as lyrical, actions that are cruel, remarks that are witty or telling, historical events that

---

[2] In the case of finite-state machines, these are Mealy machines (or—only slightly different, but basically equivalent—Moore machines).

illustrate the superiority of liberal political arrangements, and so on. And the responsively elicitable performances could include anything one had the ability to do: painting well-composed pictures, toeing the party line, riding a bike, standing the right conversational distance from someone, and so on. So the practices-or-abilities counted as 'primitive' in the sense that they are regarded as inputs to the process of algorithmic elaboration need not be crude, easily acquired, or shareable with lab rodents.

What *is* crudely behavioristic about anything specifiable as a *single-state* transducing automaton—whether it became so specifiable by a process of learning or came that way innately—is the *inflexibility* of its behavior. Though it can, by definition, respond differently to different situations, it always responds the same way to the same kind of situation. Its behavior is governed by a set of reliable differential responsive dispositions. From an algorithmic point of view, the different stimulus-kinds can be thought of as *instructions* to produce performances of the kind associated with those stimulus-kinds in the state-table that specifies the automaton. So if we are in a position to produce stimuli of desired kinds at will, we can program the SSTA to execute arbitrary *straight-schedule algorithms* over its specified behavioral repertoire. That is, we can instruct it to do any of the things it can do, to produce performances of any of the kinds in its responsive repertoire R, sequentially, in any order we like. It can follow a *list* of instructions. That is one kind of practical elaboration of basic abilities into more complex ones, but it is a poor one.

Much greater flexibility is exhibited by *finite-state transducing automata* (FSTAs). Besides responding differentially to stimuli by producing performances from its responsive repertoire, an FSTA can respond differentially by changing state (Figure 2.1). This means it can combine its primitive abilities according to *conditional branched-schedule algorithms*, which specify how it *alters* its differential responsive dispositions in response to the actual outcome of something it has done—for instance, its success or failure at achieving some recognizable result. These multi-state *functional* systems are accordingly much more capable than the corresponding single-state *behavioral* systems definable over the same set of primitive discriminable situation-kinds and elicitable performance-kinds. They can elaborate much more complex practices-or-abilities on the basis of the same primitive abilities. The advance from *behaviorism* to *functionalism* in the philosophy of mind corresponds to the move from a single-state to a multi-state
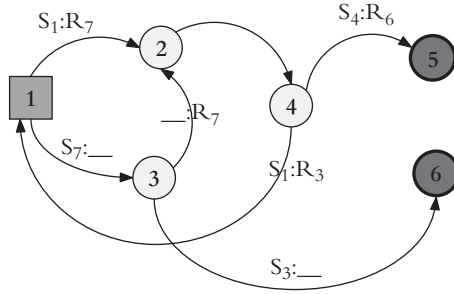
Figure 2.1 A finite-state transducing automaton

model. I will have more to say about functionalism in this connection later on.

   Transducing automata are more than merely syntactic elaborating engines because the stimuli they can respond to and the responses they can produce are not limited to symbol-types (or sign-types). Depending on the 'alphabet' of stimulus- and response-kinds they elaborate, however, they can also manipulate symbols. But they also allow us to think about symbols in a new way: still not representationally, nor yet semantically, but not just syntactically either. For we can think of *symbols* generically as anything that can both be read and written, that is, recognized and produced. In this broad sense, appropriate to transducing automata, anything in the intersection $S \cap R$ of S and R can be used as a symbol: any stimulus-kind, instances of which the system can produce as responses.

   Algorithmic elaborating abilities—paradigmatically abilities to respond differentially to stimuli by changing state—implement PP-sufficiency relations between primitive abilities and more complex ones. They accordingly delineate a clear notion of one set of practices-or-abilities being *in principle sufficient* for another, so of what it is for the elaborated practices-or-abilities to be *practically implicit in* the primitive ones from which they are algorithmically elaborated. They give a definite sense to saying that *if* a system has the abilities $A_1, \ldots, A_n$, it already knows how (is able) to do everything it needs *in principle* to know how to do in order to do A. There is a kind of idealization involved in any such claim. What is required to define the basic relation of PP-sufficiency is to be clear about the nature of that idealization. Not every system may in fact have the algorithmic elaborating abilities. Algorithmic PP-sufficiency is what holds in case *all* it needs to elaborate

its basic abilities into the complex one in question is those algorithmic abilities.

In this sense, we can diagram the meaning-use relations for two of the examples already considered so as to show the automata that implement the algorithmic elaboration PP-sufficiency relations. Figure 2.2 is an automaton-implemented, algorithmically elaborated, pragmatically mediated *syntactic* relation between vocabularies. The resultant meaning-use relation in Figure 2.3 is an automaton-implemented, algorithmically elaborated, pragmatically mediated *semantic* relation.

The automaton-theoretic metavocabulary for specifying abilities that implement PP-sufficiency relations typically involves further idealizations as well. These idealizations have not, so far as I can determine, been much discussed, but they are substantial and noteworthy in connection with the present enterprise. The first idealizing assumption is that any stimulus to which a system can respond differentially can be connected to any response
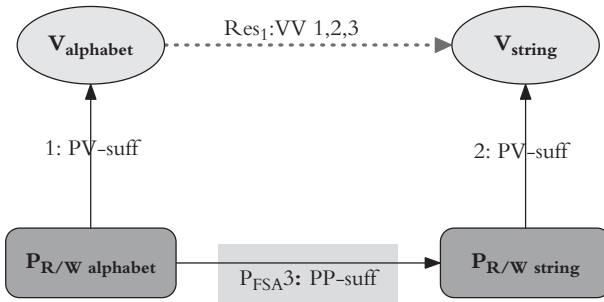

Figure 2.2 Automaton-implemented, algorithmically elaborated, pragmatically mediated syntactic relation between vocabularies
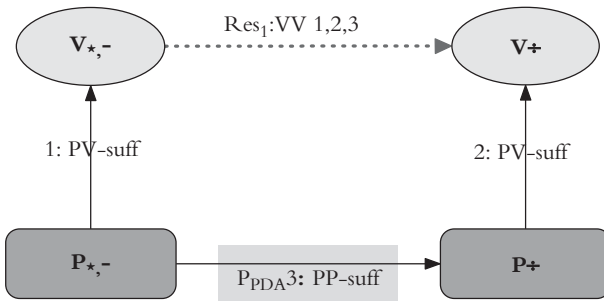

Figure 2.3 Automaton-implemented, algorithmically elaborated, pragmatically mediated semantic relation

it is capable of differentially producing. That is, there is no restriction on which elements of S and of R can be linked together as elements of the state table of an SSTA, FSTA, or PDTA. I will call this assumption "*response substitution*": if a system is capable of producing some response differentially to Ss, and can respond differentially to some kind of stimulus by producing Rs, then it is in this sense in principle capable of responding differentially to Ss by producing Rs. The capacity to discriminate Ss is available to be recruited by different responses.

For many actual systems, this is not a realistic assumption. I can respond with delight to the baby's chortle, the kitten's antics, and an unexpected glimpse of a beautiful orchid. I can distinguish disgusting smells, tastes, and sights. Could I really be trained to respond with delight to those disgusting stimuli? I am not, of course, claiming that I could. I am delineating a sense of "in-principle capability" in which I *in principle* already have all the basic discriminative and performative abilities needed for such a responsive connection. Insofar as I cannot in fact responsively hook those basic abilities together in arbitrary combinations, those restrictions count—relative to the idealization—as *psychological* restrictions, in a broad sense. (In this sense, automata have no psychology, though physical implementations of them do.)

The second idealization is that the stimulus-response connections of which a system is capable—either ideally, according to response substitution, or really psychologically—can be arbitrarily combined into *states*. If I can respond to stimuli of kind S_i with responses of kind R_j and to stimuli of kind S_k with responses of kind R_l, and if S_i and S_k are compatible and so are R_j and R_l, then I can be in a state in which I am disposed to respond to S_i with R_j *and* to S_k with R_l. I will call this idealizing assumption "*arbitrary state formation*." Once again, actual failures to form state-tables arbitrarily will be counted as due to *psychological* restrictions of the system, in the broad sense defined by contrast to this idealization.

The final idealization of the transducing-automaton model of PP-sufficiency is then "*arbitrary state permutation*," which is arbitrary formation of state-*tables*. If a system is capable of entering into responsive State$_1$ and is capable of entering into responsive State$_2$, defined over the stimulus-response-kind universe $<S,R>$, then it is capable in principle (in this sense) of responding to stimuli of any kind $S_i \in S$ by changing from State$_1$ to State$_2$ and vice versa.

The algorithmic abilities to form states as constellations of differential responsive dispositions, and to change them in response to already-discriminable stimuli, together with these idealizations concerning those abilities give a well-defined sense to the concept of one set of practices-or-abilities being PP-sufficient for another, in the sense that the latter can be exhibited as the result of *algorithmically elaborating* the former—or equivalently, that the latter can be *algorithmically decomposed or analyzed* into the latter.

3 Autonomous discursive practices, asserting, and inferring

So far, I have introduced three basic meaning-use relations (MURs), and used automaton-theoretic considerations to illustrate and clarify them. These basic MURs are:

- the PV-sufficiency of practices-or-abilities to *deploy* a vocabulary;
- the VP-sufficiency of a vocabulary to *specify* a set of practices-or-abilities; and
- the PP-sufficiency of a set of practices-or-abilities that can be *elaborated* into another, by a set of algorithmic abilities that *implement* that practical elaboration.

I have also appealed to a fourth:

- the VV-sufficiency of one vocabulary to *characterize* another (the relation of being a direct or immediate semantic or syntactic metavocabulary).

This terminology lines up with the conventions I have offered for meaning-use diagrams in the way shown in Figure 2.4.

All of these are *sufficiency* relations. Now I want to discuss briefly some corresponding *necessity* relations. In the previous lecture I mentioned one sense of pragmatic presupposition: the PP-necessity relation that obtains when it is not possible to engage in or exercise one set of practices-or-abilities unless one also engages in or exercises another.

The sense of pragmatic presupposition I want to focus on here, however, is PV-necessity: the sense in which the capacity to *say* something of a certain kind, to deploy a particular vocabulary, can require being able to
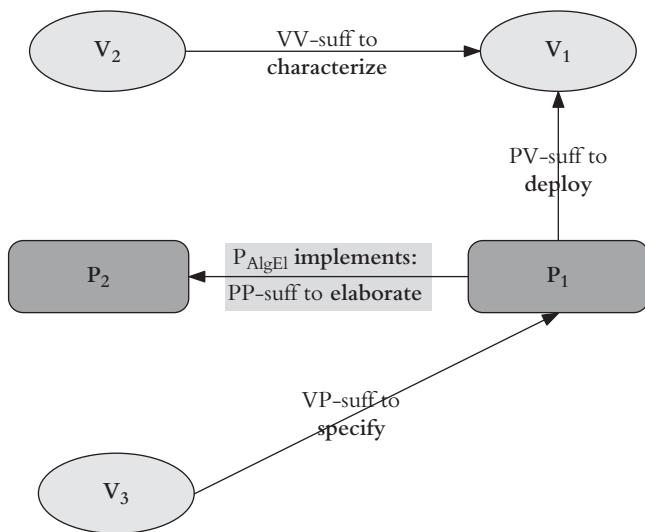
Figure 2.4 Meaning-use analysis terminology

do something of a specifiable kind. For example, using *observational* vocabulary—using expressions in the observational way, making non-inferential reports, for instance of the visible presence of red things—requires being able reliably differentially to respond to stimuli of the relevant kind, for instance visible red things. This is a capacity those who *can* deploy such vocabularies might share with those, such as pigeons, that can*not* do so.

John Dewey bequeathed to us a notion of *pragmatism* as opposed to *intellectualism* or *platonism*. These two broad currents of philosophical thought are identified with different orders of explanation: pragmatists appeal to knowing-how in order to explain knowing-that (or, more carefully, saying- or believing-that), and their intellectualist opponents (virtually the entire prior philosophical tradition, they thought) go the other way around, finding principles standing behind every propriety of practice, and rules grounding every practical ability. We are in a position to subdivide pragmatism in this traditional sense into two subsidiary claims: that for any vocabulary (any kind of saying-*that*) there are some practical abilities (some bits of know-*how*) that are PV-*necessary* to count as deploying it, and that there are some practices-or-abilities that are PV-*sufficient* to confer those contents or count as deploying that vocabulary. I have already called the latter "semantic" pragmatism. I will call the former, the PV-necessity

claim, *fundamental* pragmatism. Together they articulate the sense in which pragmatists take practices-or-abilities to be privileged with respect to, to "stand behind," the capacity to say, mean, or believe (hence to know) anything discursively.

Besides asking what practical abilities are necessary in order to deploy a *particular* vocabulary—such as observational, logical, indexical, or normative vocabulary—we can ask whether there is some kind of thing one must be able to do in order to deploy any vocabulary whatever, no matter what it is. Are there any practical abilities that are *universally* PV-necessary? This is a way of asking what sorts of *doings* deserve to count as *sayings*. In the very general way in which I have been using the term 'vocabulary', many of the vocabularies of most concern are language *fragments*: expressions whose use is not an *autonomous discursive practice* (ADP), in the sense of a language game one could play though one played no other, or a set of discursive abilities one could have though one had no other specifically discursive abilities. Singular terms, for instance, constitute a vocabulary in my sense. But no one could count as using singular terms unless she could also use sentences containing them, and hence predicates as well. Vocabularies deployed by autonomous discursive practices may be called '*autonomous*' vocabularies, or just 'languages'. (In this sense, there is no 'language of physics', though there is a *vocabulary* of physics.) Any vocabulary that is fragmentary (that is, not autonomous) pragmatically presupposes, in the PV-necessity sense, some set of autonomous discursive practices, which *are* PV-sufficient for deploying the autonomous vocabulary of which the vocabulary in question is a fragment. If that is right, then any practices PV-necessary for every *autonomous* vocabulary would be PV-necessary for every vocabulary whatsoever.

It certainly can coherently be denied that there is any core of practices-or-abilities common to all autonomous discursive practices. Wittgenstein seems to do so in thinking of such practices as language-*games* (*Sprachspiele*), given his insistence that the concept <u>game</u> does not have an essence or a definition, but is structured rather by family resemblances. That is why he does not see language as having a 'downtown', by more or less peripheral relation to which something else can count as linguistic. But I think there is a relatively bright line to be drawn in the vicinity, marking a good thing to mean by 'linguistic' or 'discursive'. Specifically, *linguistic* practices are those in which some *doings* have the practical significance of *sayings*. The

core case of *saying* something is making a *claim, asserting* something. The practices I will call 'linguistic' or 'discursive' are those in which it is possible to make assertions or claims. Although, as Wittgenstein is concerned to point out, their occurrence can have other kinds of import, the home language game of what are for that very reason called 'declarative' sentences is their free-standing use in asserting.

By this assertional pragmatic criterion of demarcation of the discursive, many of Wittgenstein's *Sprachspiele* are not really **Sprach**spiele. They are *vocal* practices, but not *verbal* ones. For instance, the practices introduced in Section 2 of the *Philosophical Investigations* are specified as

a language consisting of the words "block", "pillar", "slab", "beam". A calls them out; B brings the stone which he has learnt to bring at such-and-such a call. Conceive this as a complete primitive language.[3]

These 'calls'[4] are properly so-called. They are *signals*, appropriately responded to, according to the practice, in one way rather than another. But they are not *orders*. For an order specifies *how* it is appropriately responded to by *saying* what it is one must *do* in order to comply. "Shut the door!" can be a saying of the imperative kind only as part of a larger practice in which "The door is shut," can be a saying of the declarative kind.

I am suggesting that we treat assertional practices-or-abilities as PV-necessary for any autonomous practice we count as 'discursive'—that we think of asserting as the minimal kind of *doing* that counts as a *saying*. But what is asserting?[5] I think the beginning of wisdom in answering this question is to see that *asserting* and *inferring* are internally related practices, in the sense that each is PP-necessary for the other. Assertions are essentially, and not just accidentally, speech acts that can play the role both of premises and of conclusions of inferences. In my book *Making It Explicit*, I pursue the explanatory strategy of treating the connections in both directions also as PP-*sufficiency* relations. That is a bold and deservedly contentious theoretical approach. Here I am making only the much weaker and less controversial claim that only doings that can

---

[3] G. E. M. Anscombe (trans.), Blackwell, 1999, §2.   [4] Wittgenstein's verb is *ausrufen*.
[5] I discuss this issue at length in *Making It Explicit* (Harvard University Press, 1994). The general idea I am pursuing is laid out already in my "Asserting," *Noûs*, 17/4 (November 1983), 637–50.

serve as both kinds of *termini* for inferrings should be thought of as assertings (when the point is put pragmatically), and so that only what can stand in inferential relations should be thought of as the contents expressed by declarative sentences (when the point is put semantically). According to this way of thinking, *inferential* practices are PP-necessary components of every autonomous discursive practice, hence PV-necessary for the deployment of every autonomous vocabulary, hence PV-necessary for the deployment of every vocabulary whatsoever. They are *universally* PV-necessary.

The combination of this assertional pragmatic criterion of demarcation of the discursive and the inferential necessary condition on assertion is a distinctive kind of pragmatic *rationalism* about the discursive. *Language* games, **Sprach**spiele, must, according to this line of thought, include practices of giving and asking for *reasons*—because *assertions*, the most basic kind of *sayings*, must be capable of both serving as and standing in need of reasons. Pragmatic rationalism is the view that language *does* have a 'downtown', and it comprises the practices of making claims and giving and asking for reasons for them.

It is perhaps worth noticing in passing that this view permits a substantive response to Derrida's charge that philosophers have self-servingly fetishized reason-giving, which he sees as simply *one* game one can play with language, deserving no privilege of any kind relative to the myriad of others. Rather than simply ignoring him, or demonizing him as a dangerous irrationalist just for raising the challenge, on the one hand, or acquiescing in the radical conclusion he draws from what he takes to be the unanswerability of his challenge to justify the privileging he calls 'logocentrism', on the other, the pragmatic rationalist offers a responsive answer to that challenge: that our expressions play a suitable role in reasoning is an essential, necessary element of *our saying*, and *their meaning*, anything at all. Apart from playing such a role in justification, inference, criticism, and argument, sentences and other locutions would not have the meanings appealed to and played with by all the other games we can play with language. We philosophers should be proud to acknowledge and affirm our logocentrism, but should also justify it by an account of the relations between meaning and use, conceptual content and discursive practice.

4 Introducing conditionals

If these PV-necessity claims are correct, then there are two abilities that must be had by any system that can deploy any vocabulary, as part of the autonomous discursive practice of which the use of that vocabulary is a part: the ability to respond differentially to some sentence-tokenings as expressing *claims* the system is disposed to *assert*, and the ability to respond differentially to moves relating one set of such sentence-tokenings to another as *inferences* the system is disposed to *endorse*. These may be treated as primitive abilities for the purpose of algorithmic elaboration of further abilities, for which they are in that sense PP-sufficient. If we do that, we see that the algorithmically primitive abilities to make assertions and to sort inferences into those that are and those that are not materially good ones—in the sense of making such a distinction in practice, regardless of whether one gets the distinction *right* by some further standard—suffice for the elaboration of practices-or-abilities that are PV-sufficient to deploy a further vocabulary, namely *conditional* locutions. That constellation of basic meaning-use relations then institutes a pragmatically mediated resultant semantic relation between the original vocabulary and a more complex vocabulary that involves conditionals relating sentences deployed in the underlying ADP (Figure 2.5).

In this meaning-use diagram, the fact that the rectangle representing one set of practices-or-abilities is contained in another indicates that the smaller one is PP-necessary for the one it is contained in.
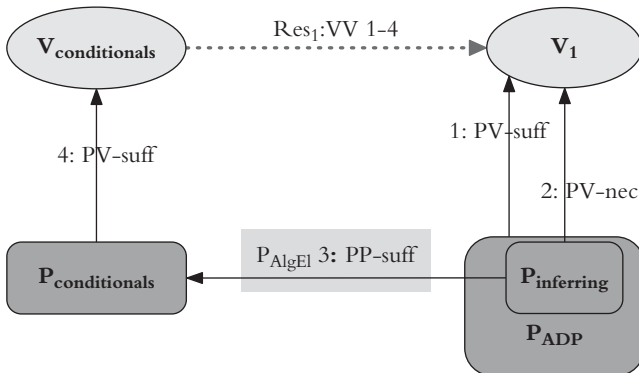

Figure 2.5 Elaborating conditionals

Here is a very rough sketch of how the algorithmic elaboration in question might work. By hypothesis, the system has the ability to respond differentially to the inference from p to q by accepting or rejecting it. It also must have the ability to produce tokenings of p and q in the form of assertings. We assume that since it can produce those assertions, we can teach it also to produce assertively tokenings of the new form "if p then q." What is required, then, is first that this new sort of response be hooked up responsively to the previously discriminable stimulus, so that it is asserted just in those cases where the inference from p to q would have been responded to as a good one. This is an exercise of the algorithmic elaborative ability I earlier called "response substitution": responsively connecting a previously distinguishable stimulus-kind to an already elicitable performance-kind. This rule codifies the *circumstances* of appropriate application of the newly introduced conditionals relating sentences deployed by the ADP. For the *consequences* of application, we need another bit of response substitution. The system can already, by hypothesis, respond to some stimuli by treating an inference as good or bad. We must now hook up that response to a new stimulus-kind. The system must respond to its *assertion* of the *conditional* "if p then q" by treating the *inference* from p to q as a good one—for instance, by being disposed to endorse q assertionally if it is disposed to endorse p assertionally. These new differential responsive abilities, achieved by reshuffling prior ones, then settle the state-table that specifies how the system is able to respond to different presented stimuli: non-logical sentences and inferences involving them, and now also conditional sentences and inferences involving them—paradigmatically, *modus ponens*. In a clear sense, then, the capacity to distinguish good from bad inferences involving *non*-logical sentences is (PP-)*sufficient* for the ability to deploy conditionals involving those sentences. That is what is represented by the MUD above.

But the pragmatically mediated semantic relation between the vocabulary of conditionals and any autonomously deployable vocabulary that obtains in virtue of the PV-necessity of material inferential practices-or-abilities for the deployment of any vocabulary includes another crucial element. For conditionals let one *say* something, where before one could only *do* something. Saying that if something is copper then it conducts electricity is a new way of doing—by saying—what one was doing before by endorsing the material inference from "That is copper" to "That conducts

electricity.'' Conditionals make *explicit* something that otherwise was *implicit* in the practical sorting of non-logical inferences into good and bad. Where before one could only in practice *take* or *treat* inferences *as* good or bad, after the algorithmic introduction of conditionals one can endorse or reject the inference by explicitly *saying* something, by asserting or denying the corresponding conditionals. What the conditional says explicitly is what one endorsed implicitly by doing what one did. The expressive role distinctive of conditionals whose use is elaborated in the way I have just specified is to *codify* inferences, to *specify* inferential practices-or-abilities, to *explicate* them, in the sense of making explicit something that was implicit in them.[6] That is to say that the MUD for the resultant MUR that is put in play by the introduction of conditionals is really the one shown in Figure 2.6.
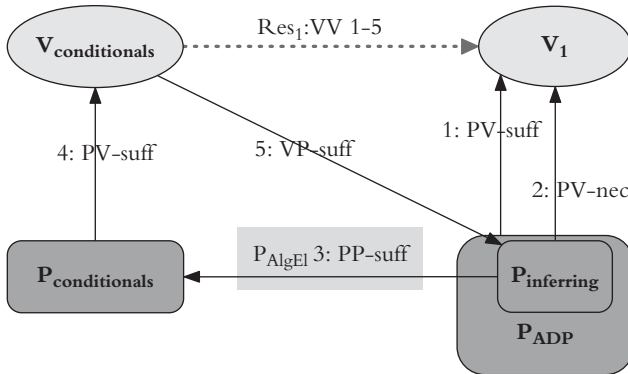

Figure 2.6 Elaborated-explicating (LX) conditionals

---

[6] The effect of asserting the new sayable that is a conditional (the practices for deploying which can be elaborated from inferential practices) is not to *say that* an act of inferring is permissible. For that one needs *normative* vocabulary, and the concepts of saying and inferring—expressed by vocabulary one need *not* master in order to master the use of conditionals. Rather, conditionals assert explicitly that one thing that can be said *follows from* another thing that can be said, that the one is a *consequence* of the other. In Lecture 4 (and further, early in 5), I discuss various sorts of semantic inferential relations among contents and their relation to pragmatic relations among deontic statuses, and say something about how to introduce the normative vocabulary that is VP-sufficient to specify this aspect of inferential practices. What I mean to be introducing here is the notion of a *genus* of relations of VP-sufficiency to express different aspects of practices PV-sufficient to deploy vocabularies. The relations between the aspects of practice made explicit by logical and modal vocabulary (those pertaining to the *content* expressed), on the one hand, and the different but complementary aspects made explicit by normative vocabulary (those pertaining to the *act* of expressing), on the other hand, are the topic of Lecture 6. Detailed consideration of the sense of ''making explicit what is implicit'' that is invoked here by the VP-sufficiency relation that constitutes the second half of the pragmatically mediated semantic relation of one's vocabulary being LX for another must accordingly await clarification until then.

Conditionals here are both *elaborated* from and *explicative* of inferential practices. As shorthand, I will call this important kind of pragmatically mediated semantic VV-relation an ''LX'' relation. It obtains when some practices-or-abilities PV-necessary for the deployment of one vocabulary can be algorithmically *elaborated* (the 'L' part) into a set that is PV-sufficient to deploy a vocabulary that is VP-sufficient *explicitly* to specify or codify the original set of practices (the 'X' part). Just in virtue of being able to assert and to sort inferences into materially good and materially bad ones (once again, whether or not correctly), one already knows how to do everything necessary in principle to introduce vocabulary that will let one *say* something, the saying of which *is* taking some inferences to be good and other ones not good.

Conditionals are a paradigm of *logical* vocabulary. For instance, the conditional is the very first bit of logical vocabulary Frege introduces in his *Begriffsschrift*, the founding document of modern logic. I want to suggest that the meaning-use analysis just offered of conditionals specifies the genus of which logical vocabulary is a species. That genus is distinguished by three characteristics:

1. being deployed by practices-or-abilities that are algorithmically elaborated from
2. practices-or-abilities that are PV-necessary for *every* autonomous vocabulary (and hence every vocabulary whatsoever) and that
3. suffice to specify explicitly those PV-necessary practices-or-abilities.

Any vocabulary meeting these conditions I will call a ''*universal* LX-vocabulary.'' In my fifth lecture I will offer more details about how one can think of other logical connectives as exhibiting this pattern of pragmatically mediated semantic relation to non-logical vocabularies generally.

By way of anticipatory illustration, I can say that I take it that just as every autonomous discursive practice must involve distinguishing some inferences as materially good, so it must involve distinguishing some claims as materially *incompatible* with others. That a monochromatic patch is red rules out its being blue. Only algorithmic elaboration is required to turn the ability to distinguish material incompatibility of claims into the ability to deploy logical *negation*. And once that bit of logical vocabulary is deployed, it (together with the conditional) lets one *say that* two claimables are materially incompatible: ''If a monochromatic patch is red, then it

is *not* blue.'' That is, negation lets one make explicit, in the form of claims—something that can be said and (so) thought—a relation that otherwise remained implicit in what one practically did, namely treat two claims as materially incompatible. So negation also qualifies as being LX (elaborated-explicitating) for every autonomous vocabulary, and hence as belonging to the pragmatically specified semantic genus within which I am suggesting we locate *logical* vocabulary.

5   Characterizing logic: the logicist's dilemma

I am suggesting that standing in this complex, resultant meaning-use relation to every autonomously deployable vocabulary can serve as a partial answer to a central question in the philosophy of logic: the demarcation question. That question is, roughly, ''What is logic?''—or, somewhat more carefully, ''What is logical vocabulary?'' That is, ''What features should be taken as distinguishing some bit of vocabulary as distinctively *logical* vocabulary?''

Even when a complete answer is envisaged as taking the form of necessary and sufficient conditions, the general demarcational question is still normally understood to concern the *circumstances* of appropriate application of the term 'logical vocabulary'. But a question of this kind comes with reasonably definite criteria of adequacy for assessing the correctness of an answer only in the context of a relatively clear specification of the *consequences* of application being associated with the expression at issue. That is, when we ask whether, say, *modal* vocabulary, or *set-theoretic* vocabulary, is a kind of *logical* vocabulary, what turns on the answer? What will we have found out about it if the answer turns out to be yes? What is the difference that *makes* a difference? The demarcational question can sensibly be addressed only if we address also the (at least co-ordinate, perhaps even prior) question concerning the theoretical, explanatory, argumentative, or constructive role logic or logical vocabulary is being envisaged as playing in some larger philosophical enterprise.

Of course, there are as many possible answers to that question as there are philosophical programs that assign some privileged role to logic. And they are liable to lead to very different answers to the demarcational question. The appeal to logical vocabulary that I want to focus on is that made by

the classical project of philosophical analysis. As I characterized that project in my first lecture, it aims to exhibit some target vocabulary as the result of semantically elaborating some base vocabulary—in the core programs I pointed to, some version of those privileged by empiricism or naturalism. I pointed out that *logical* vocabulary typically plays a special role in this enterprise: one is allowed and encouraged to use logical vocabulary in the process of elaborating one vocabulary into another. This is what I called the "semantic logicism" of the classical project of analysis. Insofar as the twentieth century analytic project did assign such a privileged status to logical vocabulary, it is fair to ask why that privilege could be thought to be legitimate. What are the rules of the analytic game? Why is it all right for the empiricist about theoretical entities to appeal to logical vocabulary, but not to modal or intentional vocabulary in elaborating his analyses? Is it all right for the naturalist to employ modal vocabulary in his elaboration of semantic vocabulary? And if so, is that because it *is* logical vocabulary? Why does logical vocabulary "come for free" in analysis, not itself having to be elaborated from the empiricist's or naturalist's basic vocabularies?

This is the larger framework in which I want to place logical vocabulary in order to get a standard for assessing various possible principles of demarcation. I want to ask the general demarcational question with the standard of assessment for answers being set by the demand that it vindicate the semantic logicist commitment of the classical project of analysis. That is, I want a way of picking out vocabulary as distinctively *logical* that explains why it is legitimate to use *logical* vocabulary, but perhaps not other kinds of vocabulary, in showing how the meanings expressed by one vocabulary can be elaborated out of the meanings expressed by another. My idea is that it is precisely whatever is an LX-vocabulary with respect to every autonomous vocabulary that can legitimately be appealed to as an auxiliary elaborating vocabulary in semantic analysis. Semantic logicism is then justified because logical vocabulary is such a universal pragmatically elaborated–explicitating vocabulary.

What conditions must a vocabulary meet in order for it to serve as a suitable auxiliary in the project of semantically analyzing one vocabulary in terms of another, in a suitably broad sense of 'analyze'? They would seem to be of two different kinds, which pull against one another so as to create what we might call the "logicist's dilemma" (looking over our shoulders

at Hempel's famous article, "The Theoretician's Dilemma,"[7] which makes a cognate point). On the one hand, the auxiliary vocabulary should not add anything distinctive of its own. For any substantive content that it contributes impugns the claim that what can be said in the terms of the target vocabulary is somehow present already in the base vocabulary, or at least is intelligible entirely in terms of what that vocabulary expresses. The general thought is that the auxiliary vocabulary can help *articulate* what is expressed in another way by the base vocabulary, can in some sense draw out what is *implicit* in it. And the claim that the version of what is expressed in the target vocabulary that emerges from that process really was already implicit in the base vocabulary, needing only further articulation to become visible as such, is undercut insofar as further new content is imported by the auxiliary vocabulary. Perhaps the semantic content in question is really implicit only in the auxiliary vocabulary, or only in the two when put together. There is no interest to the claim that culinary vocabulary supervenes, for instance, on chemical vocabulary, if it turns out we mean that it does so if we can also help ourselves to the vocabulary of home economics as an auxiliary in securing that relation. I will call this the requirement of "*semantic transparency*" on the suitability of a vocabulary to serve as an auxiliary in the analytic enterprise, that is, as playing the role there traditionally assigned to logical vocabulary.

Something like this line of thought, I think, played a significant role in twentieth-century philosophical approaches to the demarcation of logical vocabulary that appealed to its *formality*. For if logic were purely formal, concerned only with the *form* of thought and not its *content*, then logical vocabulary itself would have to be devoid of content, and hence semantically transparent in exactly the way required for it to play the role in semantic analysis that logicism assigns it. Tempting as that line of thought is, it has become increasingly difficult to maintain or pursue it.[8] For how is the contrast between semantic form and content supposed to be drawn, so as to underwrite criteria of demarcation for logical vocabulary? We are no longer in a position to subscribe to the collateral commitments required to appeal to Aristotelian or scholastic hylomorphism, nor to Kant's

---

[7] "The Theoretician's Dilemma" (1958), reprinted in *Aspects of Scientific Explanation* (Free Press, 1970).

[8] For a deep and thought-provoking extended discussion of this issue, see John MacFarlane's "What Does it Mean to Say that Logic is Formal?" (PhD dissertation, University of Pittsburgh, 2000).

transcendental faculty-hylomorphism (which are the home language-games of this distinction). And we know that a purely syntactic characterization of form, of Hilbert's sort, cannot do the semantic job we are addressing (even for the target vocabulary of arithmetic). Frege, of course, never thought of logic as purely formal, or of purely logical concepts as contentless.

The closest we come, I think, is the Frege-Bolzano identification of formality with semantic invariance under substitution, of the sort Quine has championed. To say an inference is good or a claim true in virtue of its logical form is to say two things: that it is good or true, and that it remains good or true upon arbitrary substitution of non-logical for non-logical vocabulary. This is indeed a sufficiently crisp concept to be useful for many purposes. But it is important to realize that this notion of form and formality is of no help in the context of the question about semantic logicism we are addressing—and that for two connected reasons. Notice, first, that it does not *provide*, but rather *presupposes*, a criterion of demarcation of logical vocabulary. Unless we can *already* distinguish logical from non-logical vocabulary, we cannot begin to apply the test. And that means, as Frege himself insisted, that the fact that an *inference* is *good* or a *claim true* in virtue of its logical form in this sense is not at all incompatible with its substantive contentfulness. For we can pick any vocabulary we like to privilege substitutionally: an inference is good and a claim true in virtue of its *theological* or *geological* form just in case it is good or true and remains so under all substitutions of *non*-theological for *non*-theological vocabulary, or *non*-geological for *non*-geological vocabulary. Theological and geological formality will not just depend upon, but will express an important aspect of, the *content* of theological and geological concepts. Because the substititional notion of formality is promiscuous about the vocabularies it applies to, grinding out a notion of formality for each, it leaves untouched our original problem: how to justify the claim that *logical* vocabulary, however demarcated, can play the role of expressive auxiliary in the philosophical project of wholesale semantic analysis of some vocabularies in terms of others, at least insofar as it satisfies the condition of semantic transparency.

The criterion of adequacy complementary to semantic transparency is that a candidate expressive auxiliary vocabulary in such a semantic project must be *analytically efficacious*. That is, using it must *help* in the process of establishing the desired semantic relation between vocabularies—whether

that is definability, translatability, reducibility, supervenience, or whatever. After all, it was to be the powerful modern methods of Frege's *Grundgesetze* and Russell and Whitehead's *Principia* that, within one core program of analysis, were to make the difference between *traditional* empiricism and its twentieth century successor in *logical* empiricism. For now we had much more powerful 'glue' available to stick together and articulate what is expressed by the favored base vocabularies, be they phenomenological, secondary-quality, or observational. And, looking ahead, after the subsequent *modal* logical revolution in the last third of the century, it is the power of modal vocabulary to serve as a mediating expressive auxiliary vocabulary that some philosophers hope will usher in the new age in semantic naturalism.

The logicist's dilemma—or perhaps we should say, challenge—is to explain how logical vocabulary (or any vocabulary) can be at once semantically transparent and analytically efficacious: how it can remain sufficiently semantically modest and unassuming to be eligible for use as an analytic auxiliary, while still being in a position to make a substantive contribution to the analytic semantic expressive enterprise. The potential tension between these complementary demands becomes evident when we reflect that even if we did have available a notion of the mere formality of logic that could satisfy us as to the transparency requirement, it would still be very hard to see how to understand its efficacy—how contentless concepts could help us understand contentful ones. I think the attempt to reconcile some version of these two demands has been the motive force behind a fair amount of wriggling in the philosophy of logic over the past hundred years or so. (One might think here about Wittgenstein's early agonizing about the "purity" of logic.)

The claim I want to defend is that being an LX-vocabulary with respect to every autonomous vocabulary is at least sufficient, and may be necessary, for playing the privileged role logical vocabulary is called on to play in the classical project of semantic analysis. I have been saying that showing that requires simultaneously satisfying two principal criteria of adequacy. For the first, the fact that practices sufficient to deploy logical vocabulary can be algorithmically elaborated from practices necessary to deploy any autonomous vocabulary vindicates the *semantic transparency* of logical vocabulary. For that means that anyone who can talk at all, hence can deploy any base vocabulary, can already *do* everything one needs to be

able to *do* in order in principle to be able to *say* what logical vocabulary lets one *say*. The capacity to deploy logical vocabulary (or any universally LX-vocabulary) is in this sense always already *implicit* in the capacity to deploy any vocabulary at all that might be chosen to serve as the base vocabulary of a semantic analysis or explication of any target vocabulary (whether those appropriate to empiricist, naturalist, functionalist, or any other sort of analysis). And the notion of <u>algorithmic elaboration</u> gives a definite sense to the claim that the one set of abilities is *in principle* sufficient for the other. This is the sense in which deploying logical vocabulary requires nothing new on the part of discursive practitioners: anyone who can use any base vocabulary already knows how to do everything needed to deploy any universal LX-vocabulary.

For the other criterion of adequacy, the fact that logical vocabulary makes explicit features of practices PV-necessary to deploy any autonomous vocabulary shows why and how logical vocabulary satisfies the condition of *analytic efficacy*. Here the main point is that the task for which logical vocabulary must prove itself efficacious is an *expressive* task: to show how to *say* in a *different* vocabulary what can already be said using the target vocabulary. Logical vocabulary must make it possible to *say* something one could not *say* without it. Actually introducing this explicating vocabulary and applying it (*using* the expressive resources that turn out to be implicit already in the capacity to deploy other vocabularies) is what one must *do* in order to do that: when things go well, to say in the base vocabulary, together with the logical, everything one could already say using the target vocabulary. Logical vocabulary gives one the capacity to *talk about* the inferential articulation of any other vocabulary—about what follows from what—which is an essential part of what makes that other vocabulary express the semantic content that it does. On this account, semantic logicism is a commitment to the effect that being able to make explicit the relations of material inference and incompatibility that are implicit in the use of target and base vocabularies is a critical element in establishing any relation between them that could count as a semantic *analysis* or *explication* of one in terms of the other—according to whatever more specific model of that one applies: translation, reduction, etc. It is to this *expressive* project that the *explicitating* feature of logical vocabulary (the fact that, as a universally LX-vocabulary, it is VP-sufficient to *specify* some practices PV-necessary to deploy any vocabulary) makes its distinctive contribution.

Here is a way of thinking about this vindication of semantic logicism in terms of the fact that (as I have claimed) logical vocabulary stands to any autonomous vocabulary in the complex, pragmatically mediated semantic relation of being both elaborated from and explicative of practices necessary to deploy that vocabulary. The logicist's dilemma arises if we think in purely *semantic* terms, because the first constraint (transparency) seems to require logical vocabulary to be *contentless*, while the second (efficacy) seems to require it to be *contentful*. Put that way, the conflict between the two demands is going to be hard to resolve. But if we transpose the issue into a pragmatic key, and look at what one must be able to *do* in order to *say* various things (deploy an autonomous vocabulary, deploy logical vocabulary), we see that the resolution requires only the distinction between potentiality and actuality. Transparency arises from the pragmatic fact that one is already in principle *able* to do everything one must do to deploy logical vocabulary, just by being able to talk at all. That this *pragmatic potential* was always already in place is the sense in which "nothing is added" by logical vocabulary. On the other hand, *actualizing* that potential, by introducing and deploying, actually using and applying, logical vocabulary, *does* involve *doing* something new: something one *could in principle* have done before, but had not *actually* done. The dilemma or paradox is resolved by distinguishing between what one *could* (already) do—in a sense made definite by the notion of algorithmic elaboration—and what one *actually does*. At the pragmatic level, nothing more is needed than the conceptual apparatus Aristotle introduced at the dawn of metaphysics to resolve quite a different dilemma.

6 Conclusion

The semantic relation between LX-vocabularies and the vocabularies from which they are elaborated is an essentially *pragmatically mediated* one. If that relation really is the one that explains and justifies the utility of logical vocabulary in semantic analysis, then *all* elaboration and explication, including *semantic* elaboration and explication, is implicitly *pragmatically mediated*. For in order to explain the legitimacy of appealing to *logical* vocabulary in the semantic elaboration of one vocabulary into another, we must appeal

to the *pragmatic* elaboration of one set of *abilities* into another. If that is right, then supplementing the traditional philosophical analytical concern with relations between the *meanings* expressed by different kinds of vocabulary by worrying also about the relations between those meanings and the *use* of those vocabularies in virtue of which they express those meanings—as I recommended in my first lecture—is not so much *extending* the classical project of analysis as it is *unpacking* it, to reveal a pragmatic structure that turns out already to have been implicit in the semantic project all along. For the conclusion I have been arguing for is that it is *because* some vocabularies are universal pragmatically elaborated and explicitating vocabularies that semantic analysis in the twentieth-century logicist sense is a coherent enterprise at all. That strong claim is a central element of a further kind of pragmatism about semantic analysis: what I will call *analytic pragmatism*, whose principal tool is meaning-use analysis.

My primary concern here has been to use the metaconceptual apparatus of meaning-use analysis first to make clear and then to argue for the ubiquity and theoretical centrality of pragmatically mediated semantic relations, by appeal to the idea of universal elaborated-explicitating vocabularies. Subsequent lectures will consider other philosophically important vocabularies, especially *modal*, *normative*, and *intentional* vocabularies, which I will argue can be understood as having this same general kind of meaning-use structure. In particular, in my fourth lecture I will argue that the addition of *modal* vocabulary to the semantic logicist's toolkit—the hallmark of the modal revolution of the last third of the twentieth century—can and should be vindicated in just the same way that I have suggested for ordinary logical vocabulary—though of course the features of discursive practice from which its use can be elaborated and which it in turn expresses explicitly are different. In my third lecture, building on our discussion of practical algorithmic elaboration, I shall begin an investigation of what I will urge should be considered a third core program of the classical project of philosophical analysis: functionalism about intentionality. Here, too, I will try to show, what is primarily at issue should be understood in terms of pragmatically mediated semantic relations. In particular, I will offer a new way of thinking about the commitments involved in the program of artificial intelligence, in terms of the possibility of algorithmically elaborating a set of non-discursive practices-or-abilities into one that is PV-sufficient to deploy an autonomous vocabulary.

Appendix
Indexical vocabulary: semantics and pragmatics

In the body of this lecture I introduced the notion of *algorithmic elaboration* as a way of giving definite shape to the notion of PP-sufficiency, offered some reasons to think that it plays a role in *pragmatics* analogous to that played by *logic* in *semantics*, and suggested further that we might exploit that analogy so as to understand the privileged role that logical vocabulary plays in the analytic *semantic* enterprise in terms of the elaboration of practices-or-abilities on the side of *pragmatics*. Here I want to lay out, alongside this intimation of what we might learn about *logical* vocabulary by thinking of it as a kind of universal LX-vocabulary, a slightly more extended elaboration of a case I introduced in the first lecture, one in which PP-sufficiency relations in the form of algorithmic elaboration underwrite semantic relations that we understand much better when we think of them as pragmatically mediated: that is *indexical* vocabulary, where I claimed we can find an important example of strict pragmatic expressive bootstrapping. For although we can *show* that indexical vocabulary is not semantically reducible without remainder to non-indexical vocabulary, nonetheless purely non-indexical vocabulary *can* serve as an adequate *pragmatic* metavocabulary for indexical vocabulary. That is, one can *say* (that is, describe), in wholly non-indexical terms, everything one needs to *do* in order to *use* indexical vocabulary. Non-indexical vocabulary is VP-sufficient to specify practices-or-abilities PV-sufficient to deploy indexical vocabulary. I made this claim already in Lecture 1, but at that point did not yet have on board the theoretical machinery needed to substitute a full argument for that gesture.

Indexicals exhibit two distinctive sorts of discursive behavior. On the *semantic* side, they are token-reflexive expression types: the *content* any tokening expresses depends on the *context* in which it is produced. And on the *pragmatic* side, their use can have the special pragmatic significance of making explicit the acknowledgment of a practical commitment. Marking these two dimensions of their use by pairs of names of some of those philosophers who have taught us the most about them, I will say that that indexical vocabulary has both a Kaplan-Stalnaker semantics and an Anscombe-Perry pragmatics. These are not independent. I think we can show that any expression-types whose tokenings characteristically have the Anscombe-Perry

sort of pragmatic significance *must* have their semantic contents determined in the Kaplan-Stalnaker way.

A key semantic point we have come to realize about the relations between indexical and non-indexical vocabularies is that, for any proposed translation of some indexical expressions into non-indexical terms, it is possible to describe counterfactual circumstances in which the indexical tokening *would*, and the non-indexical candidate for semantic equivalence with it would *not*, have the pragmatic significance to which Anscombe and Perry draw our attention. And that means that the indexical expression will not be intersubstitutable with—and hence not semantically equivalent to—the non-indexical one in some modal contexts. As Perry puts it, indexicals are *essential* for the expression of some thoughts. So indexical vocabulary is not semantically reducible to or analyzable in terms of non-indexical vocabulary. I think that is right and important. But that *semantic* irreducibility is best understood in the context of various other important semantic relations between indexical and non-indexical vocabulary that are *pragmatically* mediated.

In this connection, I want to make an uncontroversial observation, and a somewhat more surprising claim. The banal observation is that the Kaplan-Stalnaker semantics of character and content can be formulated in an entirely non-indexical vocabulary, and further that that account can be transformed into non-indexically formulated rules for *using* expressions with that semantics. (This is what I say in Lecture 1.) The more interesting claim (which is necessary for my overall point) is that the non-indexically specifiable practices that are PV-sufficient for using expressions that exhibit the Kaplan-Stalnaker semantics can be *algorithmically elaborated* into—and hence are PP-sufficient for—practices of using expressions that have the Anscombe-Perry pragmatic significance of *essential* indexicals. If that is right,[9] then anyone who knows how to use *non*-indexical vocabulary already knows how to do everything necessary in principle to use *indexical* vocabulary. In spite of the semantic irreducibility of the one to the other, no one can be in the position of understanding or having mastered the use of non-indexical vocabulary without having all the abilities needed to understand or master the use of indexical vocabulary. So although indexical vocabulary exhibiting the Anscombe-Perry pragmatics can*not* be *semantically reduced* to non-indexical vocabulary, a non-indexical *pragmatic metavocabulary* for it is possible. This would be a *semantic* example of *pragmatic expressive bootstrapping*. Those are the claims that I want to clarify and at least begin to justify.

Different tokenings of the same indexical expression-type can express different contents: Your utterance of 'I' picks out a different individual from mine. But those different contents are associated systematically with features of the tokenings—with, as we say, *indices* of the unrepeatable utterances, such as speaker,

---

[9] And these really are the dimensions along which the two kinds of vocabulary are distinguished.

and time and place and even the possible world in which they are uttered. The basic idea of the Kaplan-Stalnaker semantics for indexicals is that, however we understand or represent the contents expressed by particular tokenings, the semantic interpretants of indexical expression-*types* should be understood not as such contents, but rather as *functions* from tokening-indices *to* such contents.[10] To compute the time referred to by a tokening of the temporal indexical type "last Wednesday," one must know both on what day that particular, unrepeatable utterance-event was produced and what function is associated with the repeatable type (Kaplan's 'character') "last Wednesday."

Kaplan thinks that this same semantic apparatus can be applied to demonstratives—that is, that demonstratives are a kind of indexical. To do that, one must think of *demonstrations*, which make some feature of the environment semantically salient, as indices that can be associated with utterances. I think this tempting assimilation elides a very important distinction. What I want to call 'genuine' semantic indices are features of utterances that can be read off without knowing anything about what the utterance means. Time, place, speaker, and possible world are properties of tokenings that can be settled and specified before one turns one's attention to the content expressed by those tokenings. By contrast, virtually *any* feature of an utterance can be used to make something semantically salient. Wittgenstein's discussion reminds us of how important context is in determining what one is pointing at, even in the paradigmatic case of, as it were, literal demonstration. Prior conversation, for instance, can make any property of an utterance or its environment, whether gestural or not, salience-determining. 'Demonstration' as it has to function to serve as an index is not a natural kind, but a characterization of what plays a certain kind of *semantic* role. Deciding what the demonstration is in most cases requires deciding what is demonstrated. It is not, as time, place, and speaker are, something that can be determined independently of that semantic value and then appealed to as an input from which the value could then be computed by a character-function.

Kaplan's notion of a semantic index as the input to a character-function captures one way in which features of the use of expressions can contribute to their meaning. Following Stalnaker, he thinks of this phenomenon as exhausting the relevance of pragmatics to semantics. As I have been using the term 'pragmatics', of course, the contribution of the phenomena it studies to semantics is not restricted to token-reflexive expressions such as indexicals and demonstratives. But even within that restricted realm, there is a lot more to the pragmatics of these expressions than the character-relativity of content. As I have argued

---

[10] Here I am abstracting from how Kaplan and Stalnaker combine this traditional way of thinking about indexicals with their functional, possible-worlds approach to content—which is to say, the particulars of their contribution to this tradition.

at length in chapter 7 of *Making It Explicit*, for instance, the use of indexical and demonstrative vocabulary presupposes the use of *anaphoric* vocabulary. An utterance qualifies as cognitively significant and semantically contentful only if it can serve as a premise in inferences. For that reason, securing reference requires making possible non-accidental *co*-reference. For unshared proper names and uniquely identifying definite descriptions, identity of lexical type can guarantee the semantic repeatability of tokenings that supports inferences of the form: *P* of *a*, so *Q* of *a*. What makes it possible for different speakers, at different times and places, and in different situations to draw corresponding conclusions from otherwise unrepeatable indexical and demonstrative utterances is just the possibility of picking up those references anaphorically, by using pronouns, so as to say things like: "If that had been a raccoon you heard in the garden last night, we would see its tracks in the snow there now." Anaphoric uses accordingly come as part of an indissoluble practical package along with indexical and deictic ones, which would otherwise be wholly idle semantically. The underlying relation is one of *pragmatic dependence*: a PP-necessity relation, which induces a corresponding resultant semantic relation between the vocabularies deployed by the practices-or-abilities that stand in the relation of pragmatic dependence. The meaning-use diagram for this situation is set out in Figure 2.7. (Notice that this resultant MUR is the *retraction* of the PV-sufficiency relation 1 through the other two basic MURs.)

It is clear that one can state Kaplan–Stalnaker rules for associating contents with tokenings of indexical expression-types in wholly non-indexical terms. Temporal and spatial indices can be identified by non-indexical co-ordinates based on any arbitrary origin and using any arbitrary units. Then we can state rules such as:

S   A tokening of "It is raining here now," uttered at place $<x, y, z>$ and time t, is true iff it is raining at $<x, y, z>$ at t.
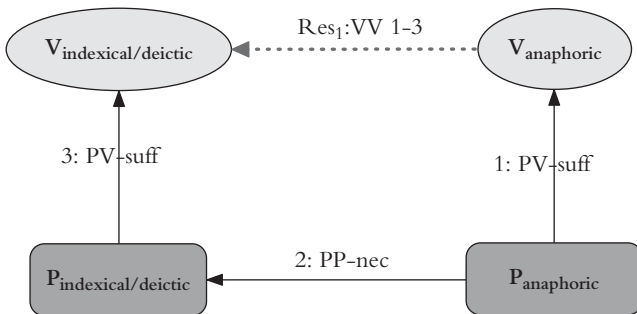

Figure 2.7 Pragmatically mediated semantic presupposition of anaphoric by indexical and deictic vocabularies

Such rules associate with each indexical character a function from non-indexically specified circumstances of utterance and circumstances of evaluation to truth values. Rules like these show that a VV-sufficient characterization of indexical vocabulary can be provided by an entirely non-indexical semantic metavocabulary. Furthermore, Kaplan and Stalnaker do that by extending the very same functional apparatus used originally to provide an extensional semantic metavocabulary for intensional vocabulary such as modal operators. And it is straightforward to transform these semantic rules into corresponding pragmatic rules for producing or assessing indexical performances. Assuming suitable units,

P   A tokening of "It was raining one mile north of here yesterday," uttered at place $<x, y, z>$ and time t, is correctly assertible iff it was raining at $<x + 1, y, z>$ at $t - 1$.

But Anscombe and Perry have shown that it does *not* follow that indexicals can be eliminated in favor of, or reduced to, non-indexical terms.[11] Although it is true (with some qualifications that are not important here) that in any utterance u containing the terms 'I', 'now', or 'here', 'I' refers to the utterer of u, 'now' refers to the time of utterance of u, and 'here' refers to the place of utterance of u, those non-indexical characterizations of the extensions of the indexical expressions are not, in general, equivalent to the indexicals. The difference lies not in their semantic extensions, but in their pragmatic significance, that is, in their *use*. Indexicals play a special role in the behavioral economies of their users. They can have a motivational significance that is not reproduced by non-indexical expressions—even those that are *de jure* co-extensional with the indexical ones. That the meeting is starting *now*, that the bear is going to eat *me*, that the treasure is buried *here* are sayings that can immediately move me to *do* things that I need not be similarly motivated to do by the realization that the meeting starts at noon, the bear is going to eat Bob, and the treasure is buried at $<x, y, z >$, even if those are the non-indexical specifications of the very indices that characterize my utterance. For any such non-indexical way of picking out the referents in question, we can come up with counterfactual circumstances in which acquiring the non-indexically expressed belief would not have the same practical significance as acquiring the indexically expressed one—and this includes candidates such as "the utterer of u," "the time of u," and "the place of u." Although it takes a somewhat *outré* scenario, it is even possible for me to acknowledge that the maker of *this very statement* has

[11] G. E. M. Anscombe, "The First Person," in Samuel Guttenplan (ed.), *Mind and Language* (Clarendon Press, 1975); John Perry, "The Problem of the Essential Indexical," *Noûs*, 13/1 (March 1979), 3–21; David Lewis, "Attitudes *De Dicto* and *De Se*," *The Philosophical Review*, 88 (1979), 513–43.

a certain commitment while not taking it that *I* do, in cases, for instance, where I do not believe that I am making that statement.

The special motivational-epistemological role that Perry points to by focusing on the sorts of practical mistakes or failures to respond appropriately to a situation that are intelligible for any non-indexically formulated beliefs, but not for some indexical ones, and that Anscombe points to by focusing on the kind of failures of identification that are *not* intelligible with some indexically formulated beliefs, but are with any non-indexical ones, is, I think, that some indexicals have *immediately commitment-acknowledging uses*. These are uses that stand in reliable causal-dispositional relations to non-linguistic responses and stimuli. Intentional (which is to say rational) agents must be able to exercise abilities reliably to respond to the acknowledgment of practical commitments whose contents we could make explicit with sentences such as "I shall raise my arm *now*," and "I shall place the cup *here*," by doing so—by "straightaway acting," as Aristotle put it. (Such practical responses are 'immediate' in the sense of 'non-inferential', since they terminate not in a saying, but in a doing.) Correspondingly, being a perceiver in the full-blooded conceptual sense of being a subject of perceptual judgments or beliefs requires the ability to respond to some environing non-linguistic situations such as the visible presence of a white cup by acknowledging doxastic commitments whose contents we could make explicit with sentences such as "There is (or I see) a white cup here now." The special pragmatic significance of some uses of indexicals to which Anscombe and Perry point is rooted in the expressive role indexicals such as 'I', 'now', and 'here' play in acknowledging practical commitments that are appropriately responded to immediately as stimuli for the production of non-linguistic states of affairs in exercises of intentional agency, and in acknowledging doxastic commitments that are appropriately elicited immediately as responses to non-linguistic states of affairs in exercises of perceptual judgment.

Of course, it is not merely accidental that it is the *same* expression-types that figure in immediately (non-inferentially) responsively acquired commitment-acknowledgments in perceptual inputs such as "The clock reads 5:05 *now*," and in immediately (non-inferentially) responded-to commitment-acknowledgments in practical outputs such as "I shall start the lecture *now*." Intentional agency requires the liability of relatively durable states of prior intention (practical commitment) that would be expressed by sentences such as "I shall start the lecture at 5:05," to *mature* into causally efficacious events of intention-in-action that would be expressed by sentences such as "I shall start the lecture *now*," which either immediately bring about or simply consist in the agent's starting to deliver the lecture. The use of the same locution 'now' that appears in the expression of the content of the intention-in-action in making explicit the non-inferentially

elicited perceptual judgment "The clock reads 5:05 *now*," makes explicit the necessary mediating role of that observation in connecting the prior intention with the intention-in-action, and so the action. Vocabulary whose expressive job essentially includes connecting perception with action so as to mediate the maturation of standing commitments into immediately practically efficacious ones will exhibit the sort of immunity to errors of misidentification of time, place, and agent Anscombe and Perry emphasize. For that connection is forged only when the very subject of a perception at a certain time and place herself acts then and there.

And for this very same reason—as a consequence of playing this same dual observational-practical expressive role—whatever expression-types do make explicit the connections between perception and agency that mediate the maturation of intentions into actions must be characterized by the Kaplan-Stalnaker semantics for expressions like 'I', 'now', and 'here'. That is, they must be what might be called 'context-homogeneous' indexicals: types whose tokenings have as their extensions at every context of evaluation whatever index characterizes their context of utterance. These contrast, for instance, with context-heterogeneous indexical types, where the time/place/person, etc., referred to is not identical to that associated with the utterance, such as 'yesterday', 'two miles north of here', and 'my wife'. These in turn are a special case of the even more general category of *cross-sortal* indexicals, such as 'my mother's favorite color', 'the dog who lives here', and 'the band that played the festival last year'. The differences between these kinds of indexicals resides entirely in *how* the character tells us to compute the content from the indices determined by the circumstances of utterance. Expression-types that mark their tokenings as available for having the pragmatic significance of expressing the immediate acknowledgment of commitments, both doxastic and practical, must function semantically as context-homogeneous Kaplan-Stalnaker indexicals because the maturation of prior intentions into actions requires that the subject, time, and place (and for that matter, world) specified in the contents of the immediately acknowledged observational and practical commitments, which engage with the content of the standing prior conditional intention, be identical with the subject, time, and place indices of the tokenings which are (when all goes well) events that are immediately causally efficacious in bringing about the intended action under the conditions envisaged. So having the Kaplan-Stalnaker semantics for context-homogeneous indexical expression-types is necessary, though not sufficient, for tokenings of those types to be lexically marked as candidates for having the distinctive Anscombe-Perry pragmatic significance.

Expressing the content of my observation as that it is 5:05 *now* is supposed to mark that in the context of a standing intention to start the lecture at 5:05,

that observation should *immediately* be responded to by the practical inten-tion-in-action whose causal efficacy is similarly marked by having its content expressed as ''I shall start the lecture *now*.'' When 'now' is used to express this role in the behavioral economy of the subject, it carries with it immunity to certain kinds of misidentification of the time at which the agent is committed to start the lecture. We might question this theoretical claim on empirical grounds. Is it not intelligible that I should acknowledge that I must start the lecture now, and still not be moved to do so? It is, but this possibility is of a different kind from that to which Anscombe and Perry point. It is a special kind of weakness of the will: *indexical akrasia*. It depends on a certain kind of breakdown in the most basic mechanisms of agency: those that connect acknowledgment of a commitment to do something to doing it, those that permit the maturation of intentions into actions. Such a breakdown is intelligible, but it is a failure of rational agency. Anscombe's and Perry's observations concern mistakes that are possible even for fully rational agents, who may fail to start the lecture at 5:05 because they do not realize that it is *then* 5:05, may not respond to threats to or opportunities for the speaker because they do not realize that *they themselves* are the speaker, and so on.

We can also ask the converse question: can there be tokenings of *non*-in-dexical expression-types—that is, ones that do *not* have the Kaplan-Stalnaker semantics characteristic of context-homogeneous indexicals such as 'I', 'now', and 'here'—that *do* have the pragmatic significance of immediate acknowledgments of practical and doxastic commitments? It may seem that this question has already been asked, and answered in the negative. But what I argued for above was the claim that if there are expression-*types* that mark their tokenings as candidates for expressing immediate acknowledgment of commitments, they must have the semantics of context-homogeneous indexicals. The current question is whether there must be such types of expression. This is a subtle and difficult question, but I think the right answer is that there *need not* be. Surely, any autonomous discursive practice must make possible the undertaking of standing practical commitments, which must be liable to maturing into intentional doings should the conditions articulated in their contents be observed to be fulfilled. Absent that possibility, practitioners would not be intelligible as agents, and hence not as speakers. Furthermore, it must be possible for actions to be controlled, and not just ballistic (''fire and forget''). That is, it must be possible for agents at least sometimes to adjust what they are doing on the basis of assessments of how successful current attempts are at reaching the desired goal. In automaton-theoretic terms, even so simple an action as reaching for a doorknob must be specified as a Test-Operate-Test-Exit (TOTE) cycle, in which each incremental movement is observed, checked against its approach to
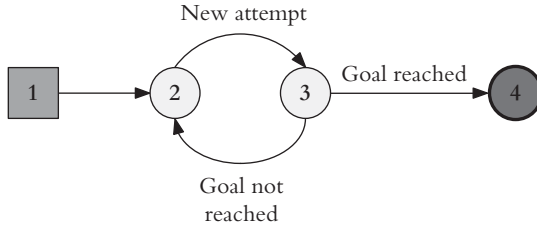
Figure 2.8 Test–Operate–Test–Exit (TOTE) cycle

the goal, and then followed by another movement calibrated by the results of the prior one, until the goal is reached (Figure 2.8).

Observation and action both require the immediate acknowledgment of commitments: acknowledgments that can cause non-linguistic performances practically, and be caused by non-linguistic events perceptually. The maturation 'here and now' of prior, standing intentions into intentions-in-action, which are practically efficacious, requires that the contents of the commitments acknowledged doxastically in perception and practically in action can sometimes coincide.

But it does *not* follow from this requirement that every discursive practice must include repeatable lexical *types*, an essential part of the expressive role of which is to mark their tokenings as fit to express the contents of those immediate commitment-acknowledgements that mediate the cycle of perception and action.[12] It is perfectly intelligible that, in some linguistic community, agents with a commitment to begin the lecture at 5:05 should be able to respond to clocks by reporting "The clock reads 5:05" and straightaway beginning the lecture. *We* would be right to describe such a person as having a 'now' belief, and in expressing the content of the intention-in-action that informs what he does by the use of 'now', even if he would not express it that way. For him to be able to act, it is enough that he can respond to the observed fulfillment of the conditions of the prior intention by acting when they are fulfilled, even if he only has available non-indexical specifications of them. Considering a related point may make the situation clearer. Gareth Evans argues convincingly that the capacity to navigate and re-identify physical objects in a spatial environment requires the practical ability to map egocentric space on to public space. It is natural for us to describe this ability in terms of mappings of indexical on to non-indexical specifications of objects and places. But Frisbee-catching dogs and successful predators evidently display the underlying abilities without being able to deploy indexical vocabulary.

---

[12] Compare this to the possibility of there being token-token identities (whether ontological or semantic) between vocabularies that are not the result of type-type identities.

Consider, then, a community that accords some tokenings of non-indexical types the pragmatic significance of immediate acknowledgements of commitments, both practical and doxastic, but that lacks expression-types that mark their token-ings as distinctively liable to play that role. Its practitioners have, in principle, the capacity to use expressions with the Kaplan-Stalnaker semantics of context-homogeneous token-reflexives. For we have seen that one can specify rules for doing so entirely in non-indexical vocabulary. The capacity to use expressions in these ways is intelligible independently of the capacity to use those expressions to mark the immediate acknowledgment of commitments. That is the central point Anscombe makes by describing a community whose members use tokens of the type 'A' token-reflexively, each to refer exclusively to him- or herself, but without according such uses the pragmatic significance characteristic of our use of 'I'. There is no way to *derive* that pragmatic significance just from the semantics of context-homogeneous indexicals. Nonetheless, in virtue of their capacity to accord use tokens of non-indexical types as immediately commitment-acknowledging, practitioners who in addition had learned to use some expression-types according to the rules for context-homogeneous Kaplan-Stalnaker characters-and-contents know how to do everything necessary in principle to use those latter expres-sions *as* pragmatically immediately commitment-acknowledging. That is, the abilities they already have can be pragmatically elaborated into the ability to use pragmatically *essential* indexicals: context-homogeneous indexicals with the full-blooded pragmatic significance of immediately acknowledging practical and doxastic commitments.

Indeed, those primitive abilities can be *algorithmically* elaborated into the ability to use 'I', 'now', and 'here' with *both* the Kaplan-Stalnaker semantics *and* the Anscombe-Perry pragmatics characteristic of *our* uses of those indexicals. The principal move, for instance, required for Anscombe's 'A' users to become 'I' users in the sense she cares about, is an exercise of the algorithmically elaborating ability I earlier called '*response substitution*'. Where one would have responded with a non-indexical immediate commitment-acknowledgment in perception, one must be able to respond by producing a different performance, but of a kind one is already capable of: producing a token of a Kaplan-Stalnaker context-homogeneous indexical expression-type. And where before one responded *to* a non-indexically expressed immediate commitment-acknowledgement by acting, now one must produce that response instead to something one could already respond differentially to, namely the use of a Kaplan-Stalnaker context-homogeneous indexical (in each case, one that meshes appropriately with standing practical commitments). These new differential responsive dispositions will then have to be recruited and integrated into the states one is already capable of entering in response to the acquisition of

a prior intention (a standing practical commitment), by arbitrary state formation and state permutation.[13] My claim is that those algorithmic-elaborating abilities are *all* that is needed to turn the capacity immediately to acknowledge practical and doxastic commitments involving non-indexicals and the non-indexically specifiable ability to use expressions with context-homogeneous semantic characters into the ability to use indexicals with *both* the semantics and the pragmatic significance characteristic of Anscombe and Perry's *essential* indexicals.

The controversial claim here is that it is intelligible that a genuinely discursive community, which must accord some tokenings the significance of being immediate acknowledgments of practical and doxastic commitments, might nonetheless lack expression-*types* that lexically mark the liability of their tokenings to have that significance. I take it to be clear that *if* there could be such a discursive community, *then* we can see how its practices-and-abilities are in principle sufficient—via algorithmic practical elaboration—for the introduction of expression-types with the pragmatics, and *therefore* the semantics, characteristic of context-homogeneous indexicals. If that is right, then such indexical vocabulary stands to the use of non-indexical vocabulary as pragmatically *elaborated* and semantically *explicitating*. That is, it stands to non-indexical vocabulary in the same LX pragmatically mediated semantic relation that conditionals stand in to non-logical sentences related by proprieties of material inference. For though the observation that the clock reads 5:05 might (in the context of a background commitment to start the lecture at 5:05) implicitly have the pragmatic significance of acknowledging a commitment to start the lecture, expressing it as the observation that it is 5:05 *now* in that context explicitly acknowledges the commitment to start the lecture *now*. For that is, as the mere commitment to start the lecture at 5:05 is not, the form in which intentions-*in-action* are expressed. There can be such intentions without a mode of expression dedicated to marking them out, but, when they are available, it is context-homogeneous *indexical* expression-types that make explicit these causally efficacious-and-elicitable acknowledgements of practical commitments (Figure 2.9).

Here we have a live and significant semantic example, for a vocabulary of independent philosophical interest, of pragmatic expressive bootstrapping. Anscombe and Perry show that the full expressive power of indexical vocabulary cannot be duplicated with non-indexical vocabulary. No non-indexical expression will behave *semantically* the way 'I', 'now', and 'here' do in sentential contexts that include operators that make explicit the situations their pragmatic observations point to: operators such as "It is possible that S rationally believes that ... but

---

[13] I have laid out in greater detail how I think this process might work for the particular case of Anscombe's 'A' language in section V. 2 of chapter 8 of *Making It Explicit*.
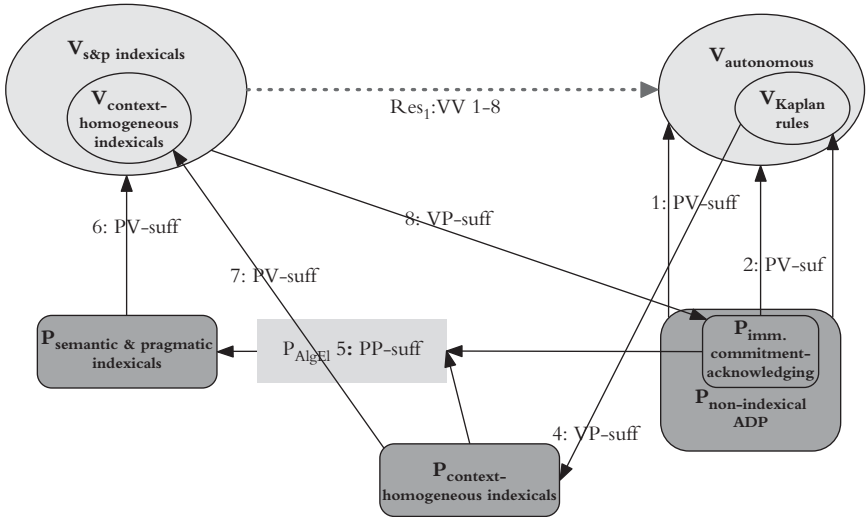
Figure 2.9 Indexicals as elaborated-explicating vocabulary

not that ... '' But in spite of this *semantic* irreducibility, non-indexical vocabulary is *pragmatically* sufficient for the *use* indexical vocabulary. The Kaplan-Stalnaker semantics permits the formulation in non-indexical vocabulary of rules VP-sufficient to specify how to use expressions as context-homogeneous indexicals, and those capacities, together with the ability immediately to acknowledge practical and doxastic commitments whose contents are characterized in non-indexical terms, can be pragmatically elaborated—using only *algorithmic* elaborating abilities (principally response substitution)—into the capacity to use indexicals in the full-blooded sense. The intermediate appeal to the Kaplan rules as a pragmatic metavocabulary makes this a more complicated case than that of conditional introduction, but what results is recognizably an instance of the form of pragmatically mediated semantic relation between vocabularies I have called 'LX'.

Why does this matter? What does this meaning-use analysis tell us? It shows us that, and how, anyone who knows how to use non-indexical vocabulary already knows how to do everything she needs to, in principle (a qualification we can cash out *precisely*, in terms of algorithmic elaborative abilities), to deploy indexical vocabulary. So one could never be in the position of understanding non-indexical vocabulary but being mystified by indexical vocabulary. Even though what is expressed by the latter cannot be fully expressed by the former, there is nothing about the use of indexicals that must remain hidden from the view of one who professes to know his way around only the non-indexical fragment of

a language. And it seems to me that the interest of this result is not hostage to the most controversial claim I made along the way. For even if the claim that there could be an autonomous discursive practice that does not include the use of indexical expression-types[14] is incorrect, I have indicated how the non-indexical *fragment* of an autonomous discursive practice suffices, as a pragmatic metalanguage (the Kaplan rules for using context-homogeneous indexicals) and via algorithmic elaboration (of non-indexical, implicitly immediately commitment-acknowledging uses into indexical, explicitly immediately commitment-acknowledging ones), for the full-blooded use of indexicals.

It may be that no one actually boggles at indexical vocabulary in a way that would make this analysis dispositive of an urgent problem rather than just generally enlightening. Indexical skepticism is certainly not widespread (and it is hard for me to see how this account will help those theologians who really are worried about the apparent inability of a non-spatiotemporal deity to think indexical thoughts). The analytic apparatus I introduced in the first lecture, and developed and applied to a live example in this one, can be thought of as ''tooling up'' for the discussion of *modal* and *normative* vocabularies in the next couple of lectures: cases where both empiricists and naturalists have had good cause for genuine puzzlement.

---

[14] Notice, however, that I did not extend that claim to *demonstratives*, which I urge should be put in a different semantic box from indexicals. Including demonstratives in the fragment of a language counted as non-indexical is not cheating in this argumentative context, for two reasons. First, the Anscombe-Perry phenomena show that indexicals are not reducible to demonstratives either. 'I' does *not* mean ''the utterer of this very sentence,'' or anything in the vicinity, as can be seen when we look at the very sort of counterfactuals they consider. Second, I did not appeal to demonstratives in pragmatically reconstructing the use of indexicals from non-indexical ingredients.

3

Artificial Intelligence and Analytic Pragmatism

1 AI-functionalism

The thought with which I introduced meaning-use analysis, and the paradigm of a pragmatically mediated semantic relation, arises when we put together two sorts of story:

- an account of what one must *do* in order to count as *saying* something—that is, of some practices-or-abilities that are PV-sufficient to deploy a vocabulary, and
- a characterization of another vocabulary that one can use to *say* what it is one must *do* to be doing something, for instance, in order to be *saying* something else—that is, of a vocabulary that is VP-sufficient to *specify* the practices-or-abilities, which might be PV-sufficient to deploy another vocabulary.

When we compose these, the resultant meaning-use relation (MUR) is the relationship between vocabularies that I have called the "pragmatic metavocabulary" relation. I have suggested that this relation is most illuminating when the pragmatic metavocabulary is demonstrably *expressively weaker* than the vocabulary for which it is a pragmatic metavocabulary. This is what I have called "pragmatic expressive bootstrapping," in the strict sense. We have seen several examples of this phenomenon:

- *Syntactic* pragmatic bootstrapping, within the Chomsky hierarchy of grammars and automata, in that expressively weaker *context-free* vocabularies are VP-sufficient to specify Turing machines (two-stack pushdown automata), which are in turn PV-sufficient to deploy (produce and recognize) expressively stronger *recursively enumerable* vocabularies.

- I argued that *non*-indexical vocabulary is VP-sufficient to specify practices PV-sufficient to deploy *indexical* vocabulary.
- I have mentioned, though not discussed, Huw Price's pragmatic naturalism, which denies the semantic reducibility of normative to naturalistic vocabulary—and even the supervenience of the one on the other—but which seeks to lessen the sting of that denial by specifying in a naturalistic vocabulary what one must *do* in order to deploy various irreducibly non-naturalistic vocabularies, for example normative or intentional ones.

I will argue in later lectures that deontic *normative* vocabulary is a sufficient pragmatic metavocabulary for alethic *modal* vocabulary: a case where the expressive ranges are at least impressively *different*, even if not rankable as strictly expressively weaker and stronger.

In this lecture, I will discuss another philosophically significant contention of this kind: the claim, thesis, or program that is usually associated with the rubric "artificial intelligence." Very crudely, AI is the claim that a computer could in principle *do* what is needed to deploy an autonomous vocabulary, that is, in this strong sense, to *say* something. It is accordingly a thesis about meaning-use relations, in my sense. The classical Turing test for the sort of 'intelligence' at issue is a *talking* test; something passes it if, by talking to it, one cannot tell it from a human speaker, that is, from someone who engages in autonomous discursive practices, someone who deploys an autonomous vocabulary. 'Intelligence' in this sense just consists in deploying such a vocabulary. Classical AI-functionalism is the claim that there is some computer program (some algorithm) such that anything that runs that program (executes that algorithm) can pass the Turing test, that is, can deploy a vocabulary in the sense in which any other language-users do. And that is to say that a computer language, in which any such algorithm can be expressed, is in principle VP-sufficient to specify abilities that are PV-sufficient to deploy an autonomous vocabulary. **So in my terms, classical AI-functionalism claims that computer languages are in principle sufficient *pragmatic metavocabularies* for some autonomous vocabulary**. (Did you see that coming?)

Now I take it that computer languages are not themselves autonomous vocabularies. For such context-free languages lack essential kinds of vocabulary. We cannot make sense of linguistic communities that speak

only Prolog or C++ (though some groups of engineers, when talking among themselves, on occasion seem to come close). Insofar as that is right, the basic claim of AI-functionalism is an *expressive bootstrapping* claim about computer languages as pragmatic metavocabularies for much more expressively powerful vocabularies, namely natural languages. Of course, AI has not traditionally been thought of as an expressive bootstrapping claim about a pragmatic metavocabulary. How could it have been? But it deserves a prominent place on the list of philosophically significant pragmatic expressive bootstrapping claims I just offered. And it should be a principal topic of philosophical meaning-use analysis. So let us see what the meaning-use analysis metavocabulary I have been deploying can help us understand about it—what lessons these metaconceptual tools can teach us when they are applied to this issue of independent interest.

Although its twentieth-century version developed later than the others, functionalism in the philosophy of mind, including its central computational species, deserves to be thought of as a *third* core program of the classical project of philosophical analysis, alongside empiricism and naturalism. (For reasons indicated in the previous lecture, I think of behaviorism as a larval stage of functionalism.) And since AI-functionalism concerns the relation between practices-or-abilities and the deployment of vocabularies, insofar as functionalist successors to behaviorist programs in the philosophy of mind do deserve a prominent place at the analytic table, that fact indicates that the sort of broadening of the analytic semantic project to include pragmatics that I have been recommending has in fact implicitly been under way for some time.

2 Classic symbolic artificial intelligence

I take the working-out of various forms of functionalism in the philosophy of mind to have been one of the cardinal achievements of Anglophone philosophy in the second half of the twentieth century. One of the things I think we have found out along the way is that functionalism is a more promising explanatory strategy when addressed to *sapience* than when addressed to *sentience*—when it is addressed to our understanding of states such as belief, rather than pains or sensations of red. In broadest terms, the basic idea of functionalism is to assimilate bits of intentional vocabulary

such as "belief that *p*" to terms classifying something in terms of the role it plays in a more complex system. So the relations between 'belief', 'desire', 'intention', and 'action' might be modeled on the relations between 'valve', 'fluid', 'pump', and 'filter'. The most immediate attraction of such an approach is the *via media* it provides between the traditional alternatives of materialism and dualism. All valves, that is, all things playing the functional role of a valve in any system, are physical objects, and they can function as valves only in virtue of their physical properties. So far, *materialism* was right: functional vocabulary applies exclusively to physical objects. But what valves have in common that makes that term properly apply to them is not a *physical* property. Mechanical hydraulic valves, heart valves, and electronic valves may have no physical properties in common that they do not share with a host of non-valves. So far, *dualism* was right: functional properties are not physical properties. *Automaton* functionalism is a species of this general view that looks specifically at the functional roles items can play in multi-state transducing automata. By the term 'AI-functionalism' I shall mean automaton functionalism about *sapience*—about what it is in virtue of which *intentional-state* vocabulary such as "believes that" is applicable to something, that is, in the terms I have been using (and which are endorsed by appeals to the Turing test), the capacity to engage in any autonomous discursive practice, to deploy any autonomous vocabulary, to engage in any discursive practice one could engage in though one engaged in no other.

So understood, AI-functionalism admits of different interpretations. Approaching it as asserting a particular kind of pragmatically mediated semantic relation between vocabularies—as making an expressive bootstrapping claim about a particular kind of pragmatic metavocabulary for some autonomous vocabulary—as meaning-use analysis suggests, leads to a characterization that is in important ways broader than traditional formulations. I want to begin by saying something about that difference.

I will call what I take to be the received understanding of the central claims of AI, what John Searle calls the "strong thesis of AI," "*classical symbolic AI*"—or sometimes "*classy* AI," for short. Here is how I understand it. Its slogan is: "*Mind is to brain as software is to hardware.*" It sees a crucial difference between modeling the mind on computer programs and all previous fashionable, rashly enthusiastic claims that some bit of impressively powerful new technology would also, *inter alia*, give us the key to unlock the secrets of the mind—telephone switchboards, clockworks, and, if we go far enough

back, even potters' wheels having been taken to play that role. For computing is manipulating symbols according to definite rules (the algorithms implicit in automaton state-tables). And, the claim is, thinking or reasoning, the fundamental sort of operation or activity that constitutes sapience, just *is* manipulating symbols according to definite rules. This *computational theory of the mind* is the basis of the standard argument for AI–functionalism. It is a view that long antedates the advent of computers, having been epitomized already by Hobbes in his claim that "reasoning is but reckoning."

Now the plausibility of understanding thinking as symbol-manipulation at all depends on taking symbols to be more than just sign-designs with a syntax. They must be *meaningful, semantically* contentful signs, whose proper manipulation—what it is *correct* to do with them—depends on the meanings they express, or on what they represent. Traditionally, this fact meant that there was a problem reconciling the computational view of the mind with naturalism. Physics does not find meanings or semantic properties in its catalogue of the furniture of the world. They are not, or at any rate not evidently, physical properties. So how could any physical system *be* a computer—a *symbol*-manipulator in the relevant sense—and so respond differentially to signs depending on the *meanings* they express? Looking back from the vantage point vouchsafed us by the development of actual computing machinery—and the realization that doing numerical calculation by the algorithmic manipulation of numerals was only one instance of a more general symbol–manipulating capacity—provides a possible answer. Already for Descartes, the thoroughgoing isomorphism he had established between algebraic formulae and geometric figures suggested that manipulating the formulae according to the rules proper to them could not just *express*, but also *constitute* or *embody*, an understanding of the figures. The isomorphism amounts to an encoding of *semantic* properties in *syntactic* ones. A physical system can accordingly be a computer—manipulate symbols in ways that accord with their meanings—because such an encoding ensures that, in Haugeland's slogan, if the automaton takes care of the *syntax*, the *semantics* will take care of itself.[1]

Usually, though, what you get when you manipulate symbols in ways that exploit isomorphisms to what they are symbols of is a *simulation*. Computers

---

[1] This characterization of classical symbolic AI owes a lot to John Haugeland's *Artificial Iintelligence: The Very Idea* (MIT Press, 1989). My own thoughts on this subject were worked out in the course of teaching undergraduate AI courses based on this text.

can manipulate symbols to model traffic patterns, weather systems, and forest fires. No one is liable to confuse the symbol-manipulating with the phenomena it simulates—the computation with the traffic, the weather, or the fire. But AI-functionalism claims that, unlike these cases, manipulating symbols in ways that suitably respect, reflect, and exploit isomorphisms with what those symbols for that very reason count as expressing or representing is not just a *simulation* of thinking, but *is* thinking itself. That is what it *is* to deploy a vocabulary *as* a vocabulary, that is, as meaningful. The *only* reason for according thought this uniquely privileged position—as the one phenomenon that cannot be symbolically *simulated* without thereby being actually *instantiated*—is whatever reason there is to think that the symbolic-computational theory of the mind is correct. And that is a very substantive, potentially controversial theory of sapience, with a correspondingly large burden of proof.

3 A pragmatic conception of artificial intelligence

I think that symbolic AI's focus on the Turing test is appropriate. There just is no point in insisting that something that is genuinely indistinguishable (including, crucially, dispositionally counterfactually) from other discursive practitioners in conversation—no matter how extended and wide-ranging in topic—should nonetheless not be counted as *really* talking, so thinking (out loud), and deploying a meaningful vocabulary. But although the slide can seem unavoidable, it is a long way from acknowledging the criterial character of the Turing test for sapience to endorsing the computational theory of the mind on which classical symbolic AI is predicated. The line of thought I have just rehearsed invites a focus on the issue of the *symbolic* character of thought that I think is ultimately misleading. And for that reason it *mis*locates, as it seems to me, what really is the most important issue in the vicinity: the claimed *algorithmic* character (or characterizability) of thought or discursive practice.

In Lecture 2 I argued that, from the point of view of meaning-use analysis, the principal significance of automata does not lie in their capacity to manipulate symbols, but rather in their implementing a distinctive kind of PP-sufficiency relation. Multi-state transducing automata *algorithmically elaborate* a set of primitive abilities into further abilities—abilities which, just

because they can be so exhibited, can then be regarded as complex, as prag-
matically *analyzable* into those primitive abilities plus the basic algorithmic
elaborating abilities. This characterization of automata suggests that AI be
understood broadly as a claim to the effect that such an analysis or decompo-
sition is possible of some autonomous discursive practice—the practice-or-
ability to deploy some vocabulary that can be deployed though one deploys
no other. That is, it claims that some autonomous discursive practice can be
exhibited as the algorithmic elaboration of a set of primitive abilities, which
are accordingly PP-sufficient for that autonomous discursive practice.

That claim by itself would not be interesting or controversial. For the null
elaboration is also an algorithmic elaboration (albeit a degenerate one). So
the condition would be trivially satisfied, just because there *are* autonomous
discursive practices-or-abilities. What is needed to turn the claim that
some set of primitive abilities can be algorithmically elaborated so as to
be PP-sufficient for some autonomous discursive practice into a genuinely
substantive claim is a further constraint on the primitive abilities. Given the
reasons for being interested in AI-functionalism in the first place, what we
want is to stipulate that what are to be counted as primitive abilities with
respect to such an algorithmic elaboration must not themselves in some
sense already be *discursive* abilities.

Here is the version that I propose. What I will call the "algorithmic
pragmatic elaboration" version of AI-functionalism—or just "pragmatic
AI"—is the claim that there is a set of practices-or-abilities meeting two
conditions:

1. It can be algorithmically elaborated into (the ability to engage in) an
   autonomous discursive practice (ADP).
2. Every element in that set of primitive practices-or-abilities can intel-
   ligibly be understood to be engaged in, possessed, exercised, or
   exhibited by something that does *not* engage in any ADP.

In the terminology of meaning-use analysis, the first of these is a kind of PP-
sufficiency claim—specifically, an algorithmic elaboration PP-sufficiency
claim. The second is the denial of a set of PP-necessity claims.

This approach to AI-functionalism shifts the focus of attention away
from the role of *symbols* in thought, away from the question of whether
thinking just *is* manipulation of symbols, and away from the issue of
whether isomorphism is sufficient to establish genuine ('original', rather

than merely 'derivative') semantic contentfulness. It is true that I am here still thinking of what is at issue in sapience as a matter of deploying *vocabularies*, that is, using symbols, semantically significant signs—not in a derivative way, but in whatever way is fundamental in the sense of being exhibited by *autonomous* discursive practices-or-abilities, and the vocabularies they deploy. But—and here is the important difference from classical symbolic AI—the connection to computers (or as I would prefer to say, *automata*) is established not via the principle that computers are symbol-manipulating engines and that, according to the computational theory of the mind, thinking just consists in manipulating symbols, but rather via PP-sufficiency of the algorithmic elaboration sort that I discussed in Lecture 2. And the structural question AI-functionalism asks is an issue that can arise for *any* ability—not just those that involve symbol use. That is, for *any* practice-or-ability P, we can ask whether that practice-or-ability can be algorithmically *decomposed* (pragmatically analyzed) into a set of primitive practices-or-abilties such that:

1. they are PP-sufficient for P, in the sense that P can be algorithmically elaborated from them (that is, that *all* you need in principle to be able to engage in or exercise P is to be able to engage in those abilities plus the algorithmic elaborative abilities, when these are all integrated as specified by some algorithm); and

2. one could have the capacity to engage in or exercise *each* of those primitive practices-or-abilities without having the capacity to engage in or exercise the target practice-or-ability P.

If those two conditions are met, we may say that P is **substantively algorithmically decomposable** into those primitive practices-or-abilities. So, for instance, the capacity to do long division *is* substantively algorithmically decomposable, into the primitive (with respect to this decomposition) capacities to do multiplication and subtraction. For one can learn to multiply, or again, to subtract, without yet having learned how to divide. Perhaps (though I doubt it) the capacity to play the piano is like this, since one can learn how to finger each key individually, and to adjust the intervals between doing so. By contrast, the capacities to respond differentially to red things and to wiggle my index finger probably are not substantively algorithmically decomposable into more basic capacities. These are not things that I do *by* doing something else. If I do not have those abilities, there

is no way to put them together as the complex results of some structured sequence of other things—even with the flexibility of conditional branched schedule algorithms, hence of Test-Operate-Test-Exit feedback loops of perception, action, and further perception of the results of the action. The abilities to ride a bicycle, to swim, or to hang-glide might or might not be substantively practically algorithmically decomposable, and the empirical question of whether they are, and if so, how, is of considerable pedagogical significance (about which more later).

So the question of whether some practice-or-ability admits of a substantive practical algorithmic decomposition is a matter of what contingent, parochial, matter-of-factual PP-sufficiencies and necessities actually are exhibited by the creatures producing the performances in question. That question is very general and abstract, but also both empirical and important. It is a very general *structural* question about the ability in question. That issue as such, however, has *nothing whatever* to do with *symbol* manipulation. My suggestion is that we think of the core issue of AI-functionalism as being of this form. The issue is whether *whatever* capacities constitute sapience, *whatever* practices-or-abilities it involves, admit of such a substantive practical algorithmic decomposition. If we think of sapience as consisting in the capacity to deploy a vocabulary, so as being what the Turing test is a test for, then since we are thinking of sapience as a kind of symbol use, the target practices-or-abilities will also involve symbols. But that is an entirely separate, in principle independent, commitment. That is why I say that classical symbolic AI-functionalism is merely one species of the broader genus of algorithmic practical elaboration AI-functionalism, and that the central issues are mislocated if we focus on the *symbolic* nature of thought rather than the substantive practical algorithmic analyzability of whatever practices-or-abilities are sufficient for sapience.

4 Arguments against AI-functionalism: ranges of counterfactual robustness for complex relational predicates

Because the two stand or fall together, arguments against the plausibility of the claims of classic symbolic AI-functionalism usually take the form of arguments against the computational theory of the sapient mind.

These arguments include doubts about the possibility of explicitly cod-
ifying in programmable, hence explicitly statable, *rules* all the implicit
practical background skills necessary for thoughtful engagement with the
world, challenges to the adequacy of the semantic epiphenomenalism
inherent in treating syntactic isomorphism as sufficient for non-derivative
contentfulness, and reminders of the sort epitomized by Searle's Chinese
Room thought-experiment[2] of how badly the essentially third-person
point of view of this sort of functionalist successor to behaviorism fits
with intuitions derived from our first-person experience of understand-
ing, grasping meanings, deploying vocabularies, and having contentful
thoughts. Reasons for skepticism about the sort of AI understood instead as
claiming the substantive algorithmic decomposability of autonomous dis-
cursive practices-or-abilities into non-discursive ones must take a distinctly
different shape.

   For instance, Dreyfus objects to classical symbolic AI on the grounds
that it requires that all the implicit practical skills necessary for under-
standing our ordinary life-world have to be made explicit in the form of
rules (codified in programs).[3] He diagnoses classy AI as built around the
traditional platonist or intellectualist commitment to finding some bit of
explicit knowing- (or believing-)*that* behind every bit of implicit practical
knowing-*how*. Like Dewey, he is skeptical about that framing commit-
ment. By contrast, the corresponding argument against the substantive
practical algorithmic decomposability version of AI would have to offer
reasons for pessimism about the possibility of algorithmically resolving
essentially discursive knowing- (or believing-)*that* without remainder into
non-discursive forms of knowing-*how*. **Whatever problems there may
be with this kind of AI, they do not stem from some hidden
intellectualism, but, on the contrary, concern the particular variety
of *pragmatism* it articulates: algorithmic pragmatism about the dis-
cursive**. For what makes the substantive algorithmic practical elaboration
model of AI interesting is the relatively precise shape that it gives to the
pragmatist program of explaining knowing-that in terms of knowing-how:
specifying in a non-intentional, non-semantic vocabulary what it is one
must *do* in order to count as deploying some vocabulary to say something,

---

² In "Minds, Brains, and Programs" (1980), reprinted in John Haugeland (ed.), *Mind Design II* (MIT
Press, 1997).

³ For instance, in Hubert L. Dreyfus, *What Computers Still Can't Do* (MIT Press, 1997).

hence as making intentional and semantic vocabulary applicable to the performances one produces (a kind of pragmatic expressive bootstrapping).

What arguments are there against this pragmatist version of AI? The form of the claim tells us that to argue against the practical algorithmic elaboration version of AI we must find some aspect exhibited by all autonomous discursive practices that is not algorithmically decomposable into non-discursive practices-or-abilities. That would be something that is PV-necessary for deploying any autonomous vocabulary (or equivalently, PP-necessary for any ADP) that cannot be algorithmically decomposed into practices for which no ADP is PP-necessary.

I do not claim to have a knock-down argument here. But the best candidate I can think of to play that role is the practice of doxastic updating—of adjusting one's other beliefs in response to a change of belief, paradigmatically the addition of a new belief.

It is pretty clear that this set of practices-or-abilities is a PV-necessary aspect of the deployment of any vocabulary. For any set of practices to count as *discursive*, I claimed last time, it must accord some performances the significance of *claimings*. It is a necessary feature of that significance that what is expressed by those performances stands to other such contents in broadly *inferential* relations of being a reason for or against. That is, the practical significance of claiming includes undertaking a commitment that has other commitments and entitlements (or lack of entitlements) to commitments as its consequences, that can itself be a consequence of other commitments, and whose entitlement also depends on its relation to one's other commitments. One understands or grasps the content expressed by some bit of vocabulary that can be used to make claims only to the extent to which one can tell in practice (respond differentially according to) what follows from it and what it follows from, what other commitments and entitlements the various commitments it can be used to undertake include and preclude. And that is to say that one understands what a bit of vocabulary means only insofar as one knows what difference undertaking a commitment by its use would make to what else the one using it is committed or entitled to—that is, insofar as one knows how to update a set of commitments and entitlements in the light of adding one that would be expressed using that vocabulary (keeping deontic score). Discursive understanding of this sort is a more-or-less affair. One need not be omniscient about the significance of a bit of vocabulary in order to

deploy it meaningfully. But if one has *no* idea what practical consequences for other commitments a claim using it would have, then one associates no meaning with it at all.

If all that is right, then the question of whether doxastic updating can serve as a reason to be pessimistic about the practical algorithmic elaboration version of AI comes down to an assessment of the prospects for a substantive algorithmic decomposition of the ability to update. Why might one think that no such decomposition is possible—that is, that that essential discursive ability could not be algorithmically elaborated from any set of non-discursive abilities? The key point, I think, is that the updating process is highly sensitive to collateral commitments or beliefs. The significance of undertaking a new commitment (or relinquishing an old one) depends not just on the *content* of *that* commitment, but also on what *else* one is already committed to. I will argue in my next lecture that we can think of this global updating ability as a collection of sub-abilities: as the capacity, in one's actual doxastic context, to associate with each commitment a range of counterfactual robustness. To do that is to distinguish, for each commitment (including inferential commitments), which further commitments *would*, and which would *not*, infirm or defeat it. This includes not only claims that are incompatible with it, but also claims that are incompatible with it in the context of one's other collateral beliefs—that is, which complete a *set* of claims that are jointly (but perhaps not severally) incompatible with it.

I take it that there is nothing unintelligible about having such practical abilities, fallible and incomplete though they may be, to distinguish claims that *are* from those that are *not* contextually incompatible with a given claim. And it is clear that a global updating capacity can be algorithmically elaborated from such abilities to discriminate ranges of counterfactual robustness. But I do not think that this sort of ability is a good candidate for an algorithmic decomposition that is *substantive* in the sense I have given to that term. For I do not see that we can make sense of abilities to discern ranges of counterfactual robustness being exhibited, whether severally or collectively, by *non-discursive* creatures. The problem is that the productivity of language guarantees that anything that can talk can form predicates specifying an indefinitely large class of relational properties. As a consequence, any new information about any object carries with it new information of a sort about every other object. For *any* change in

any property of one changes *some* of the relational properties of *all* the rest. The problem in a nutshell is that doxastic updating for language-users requires distinguishing among all of these, those that *are* from those that are *not* relevant to the claims and inferences one endorses—that is, those which fall within the range of counterfactual robustness of those claims and inferences. And it is *not* plausible, I claim, that *this* ability can be algorithmically decomposed into abilities exhibitable by non-linguistic creatures.

Why not? The logical and computational versions of what the AI community calls the "frame problem" showed that updating requires exercising what turns out to be a crucially important but easily overlooked cognitive skill: the capacity to *ignore* some factors one is capable of attending to. But worrying about the practical engineering problem of how to implement such an ability in finite-state automata revealed a deeper theoretical conceptual problem, which concerns not *how* to ignore some considerations, but *what* to ignore. A simple version of the issue is afforded by the familiar observation that anything is similar to anything else in an infinite number of ways, and also dissimilar to it in an infinite number of ways. For instance, my left little finger and Bach's second Brandenburg concerto are not only different in countless ways, but are similar in that neither is a window-shade, nor a prime number, neither existed before 1600, and both can be damaged by the careless use of stringed instruments. Dealing with objects as knowers and agents requires the ability to *privilege* some of these respects of similarity and difference—to sort the myriad of such respects into those that *are* and those that are *not* relevant to or significant for the inferences, theoretical and practical, to and from the claims about those objects with which one is concerned. In the sort of case I want to focus on, there are lots of complex relational properties that we should usually ignore in our reasoning.

For instance, Fodor defines any particle as being a 'fridgeon' just in case his fridge is on.[4] So when his fridge turns on, it also turns all the particles in the universe temporarily into fridgeons, and gives every macroscopic physical object the new property of being made of fridgeons. Again, a death in a distant place can give me the new property of having the same eye-color as

[4] In "Modules, Frames, Fridgeons, Sleeping Dogs and the Music of the Spheres," in Z. Pylyshyn (ed.), *The Robot's Dilemma: The Frame Problem in Artificial Intelligence* (Ablex, 1987).

the oldest living inhabitant of Provo, Utah. Usually I ought to ignore these properties and facts. One of the lessons of the narrower, engineering versions of the frame problem is that updating becomes computationally infeasible if I cannot do that, and am accordingly obliged to check every one of my beliefs and the inferences that support them to see whether they are infirmed by those facts—to be sure that my conclusion that the solid floor will bear my weight is not affected by its suddenly consisting of fridgeons and that my inferential expectation that I will see better if I put on my glasses is still a good one even though my eyes have the new Provo property. For a while there was a small philosophical industry devoted to trying to distinguish what Geach (thinking of McTaggart) called 'Cambridge changes' from real ones.[5] I think we have come to see that this enterprise is a misguided one. For any complex relational property such as being a fridgeon or having old-Provo-colored eyes, we can describe *some* inferential circumstances (however *outré*) in which the credentials of some significant claim would turn precisely on the presence or absence of that property. What we need to be able to do is not to classify some properties as, in effect, irrelevant *tout court* (irrelevant to what?), but for each inference, to distinguish the considerations that are irrelevant to its goodness, which should accordingly be ignored. This ability is necessary to deal with what Fodor calls epistemological 'isotropy': the fact that any belief is potentially evidentially relevant to any other, given a suitable context of collateral beliefs.

I am claiming that:

- One cannot talk unless one can *ignore* a vast variety of considerations one is capable of attending to, in particular those that involve complex relational properties, that lie within the range of counterfactual robustness of various inferences.

- Only something that can *talk* can do that, since one cannot *ignore* what one cannot *attend* to (a PP-necessity claim), and for many complex relational properties, only those with access to the combinatorial productive resources of a *language* can pick them out and respond differentially to them. No non-linguistic creature can be concerned with fridgeons or old-Provo eye colors.

- So language use, deploying autonomous vocabularies, brings with it the need for a new kind of capacity: for each inference one entertains,

---

[5] *God and the Soul* (Routledge, 1969), 71.

to distinguish in practice among all the new complex relational properties that one comes to be able to consider, those that are, from those that are not relevant to assessing it.

- Since non-linguistic creatures have no semantic, cognitive, or practical access at all to most of the complex relational properties they would have to distinguish to assess the goodness of many material inferences, there is no reason at all to expect that that sophisticated ability to distinguish ranges of counterfactual robustness involving them could be algorithmically elaborated from the sorts of abilities those creatures do have.[6]

5 Practical elaboration by training

One might reasonably wonder whether, if the sort of argument I have sketched against the substantive algorithmic decomposability of autonomous discursive practices were successful, it would not prove too much. Non-linguistic creatures do, after all, acquire the ability to engage in discursive practices. They do cross the boundary I have been worrying about, and begin deploying vocabularies. This is true both of human infants and, at some point in the past, of our hominid ancestors. The ontogenetic and phylogenetic acquisition of discursive capacities did not, and does not, happen by magic. If discursive practices-or-abilities really are *not* substantively algorithmically decomposable without remainder into non-discursive ones, how *are* we to understand the development of discursive out of non-discursive practices?

I think the answer is that besides algorithmic elaboration there is another, more basic sort of PP-sufficiency relation—another way in which one set

---

[6] This last claim is a somewhat delicate one. I am *not* using as a premise the claim that we cannot make sense of the possibility of substantively algorithmically decomposing the capacity to be aware of a full range of complex relational properties, by deploying a suitable vocabulary. That is part of the conclusion I am arguing for. I *am* claiming, first, that the ability to ignore the vast majority of complex relational properties that are irrelevant to a given inference in the sense that they fall within its range of counterfactual robustness cannot be taken as *primitive* with respect to a *substantive* algorithmic practical decomposition of discursive practices-or-abilities, and, second, that we have no idea at all how even primitive non-discursive abilities that *could* be substantively algorithmically elaborated into the capacity to form the complex predicates in question could be further elaborated so as to permit the sorting of them into those that do and those that do not belong in the range of counterfactual robustness of a particular inference.

of practices-or-abilities can practically suffice for the acquisition of another. Sometimes those who can engage in one set of practices can learn or be trained to engage in another—not because the target practices can be *algorithmically* elaborated from the original ones, or from some further set into which they can be decomposed, but just because, as a matter of contingent empirical fact concerning creatures of that particular kind, anyone who has the one set of capacities can be brought to have the other as well. So it might be that those who can draw realistic portraits of horses can be brought also to draw realistic portraits of humans, forge signatures, fold origami gracefully, and arrange flowers. If so, no doubt our account of why these other abilities were especially accessible to those who possess the original one would invoke something like ''eye-hand co-ordination'' or ''fine-muscle control.'' But that is not at all to say that there must be some set of specifiable basic abilities out of which, say, the capacity to draw a good likeness of a friend could be *algorithmically* elaborated. That capacity might admit of no algorithmic decomposition. Certainly the fact that people who can do some other sorts of things can learn or be taught also to do this does not entail or require that there be such a decomposition.

When as a matter of fact there is a course of practical experience or training that will bring those who have one set of abilities to have another set of abilities, I will say that the second can be ''practically elaborated by training'' from the first. Like algorithmic elaboration, practical elaboration by training is a kind of PP-sufficiency relation. The hallmark of the difference between them is that we can say exactly and in advance what the practices that *implement* PP-sufficiency by algorithmic elaboration are—what *else* besides exercising the primitive abilities one must be able to do in order to elaborate them algorithmically into the target ability. These *elaborative abilities* are things like response substitution and arbitrary state formation—and in general the abilities that suffice to execute a conditional branched-schedule algorithm. These algorithmic elaborative abilities are all that is needed, for instance, to turn the capacity to multiply and subtract into the capacity to do decimal division. And we know how to build machines that have *these* elaborative abilities. By contrast, in the case of practical elaboration by training, we have no idea how to specify in advance the abilities that implement the sufficiency of rote repetition for memorizing the alphabet, or practice for catching a ball or drawing a

recognizable face. And where we *can* say something about the abilities that implement PP-sufficiency relations of the practical-elaboration-by-training sort, we find that they both vary wildly from case to case, and depend heavily on parochial biological, sociological, historical, psychological, and biographical contingencies. Finally, where the question of whether one set of well-defined practices-or-abilities can be elaborated *algorithmically* into another is one that can in principle be settled a priori, from one's armchair, the question of whether it is *practically* PP-sufficient for some particular creature or kind of creature, in a particular context, by some specified training regimen, is one that can only be settled empirically.

I think an appreciation of the centrality of this sort of PP-sufficiency relation—which obtains when, as a matter of fact, creatures of a certain sort who can engage in a practice (exhibit an ability) can be brought or can learn to engage in (or exhibit) another—is one of the master ideas animating the thought of the later Wittgenstein. Again and again he emphasizes the extent to which our discursive practices are made possible by the fact that, as a matter of contingent fact, those who have one set of abilities or can engage in one set of practices can be brought by training to exhibit or engage in another. We can be trained to count, associate sounds with written shapes, and respond to signposts, and to exercise those abilities in new cases by "going on in the same way" as others who share our training (and wiring) would. Wittgenstein is, of course, concerned to show us to what extent and in how many ways our discursive practices-or-abilities depend on things that we could not be *taught* to do (by being told) if we could not be *trained* to do them (by being shown). But I think he also sees practical elaboration by training as the principal motor of our discursive practices-or-abilities, as what gives them their *theoretically motley* but *practically tractable* shapes. As I read him, Wittgenstein thinks that *the* most fundamental discursive phenomenon is this way in which the abilities required to deploy one vocabulary can contingently be practically *extended*, elaborated, or developed so as to constitute the ability to deploy some further vocabulary. We may think in this connection of the examples I mentioned in Lecture 1, of the sort of thought-experiments he invites us to conduct concerning this sort of process of *pragmatic projection* of one practice into another: the fact that people who could already use proper names for people could catch on to the practice of using them also for rivers, and that people who could already talk about having gold in their teeth could catch

on to talking about having pains in their teeth. The way in which prior abilities are recruited by training in the service of developing new ones is in general unsystematic, not codifiable in rules or algorithms, and not predictable or explicable from first principles. Wittgenstein sees this sort of non-algorithmic practical elaboration as ubiquitous and pervasive. It results in a permanent process of practical discursive mutation that is on the one hand mediated by the productivity of language, and on the other limits its diachronic systematicity.

So the answer to the question with which I began this section is that we do not need to assume that discursive practice is substantively *algorithmically* decomposable into non-discursive practices-or-abilities, on pain of making entering into those practices and acquiring those abilities—by us as a species, and as individuals—unintelligible, because there is another sort of PP-sufficiency relation besides algorithmic elaboration: practical elaboration by training. We need to acknowledge this sort of PP-sufficiency in any case, in order to account for the provenance of the abilities treated as primitive for the purposes of algorithmic elaboration. And Wittgenstein urges us to see this sort of elaboration not only as crucial for the advent of discursive practices-or-abilities, but also as pervasive within up-and-running discursive practices, alongside algorithmic elaboration.

I said at the outset of my story that one of the aims of the sort of analytical pragmatism for which I am seeking to sketch a theoretical basis is to show how Wittgenstein's pragmatist insights need not be taken to underwrite a theoretical quietism antithetical to the project of traditional philosophical analysis, but how those insights can instead be taken on board and pressed into the service of a further pragmatic development and elaboration of that project. Acknowledging the pervasiveness and centrality of non-algorithmic practical elaboration by training need not be the death of theoretical analysis of discursive practice and its relation to the semantic contents expressed by the vocabularies deployed in that practice. For the analytical Wittgenstein-ian pragmatist, appeal to algorithmically non-decomposable, contingent, parochial abilities is compatible with investigating PP-sufficiency and PP-necessity *dependency* relations between such abilities and practices, as well as the PV- and VP-sufficiency relations they stand in to vocabularies. I would like to close this lecture by outlining one analytic issue that I think is raised directly by the consideration of what I will call *pedagogical* practical elaboration and decomposition of practices and abilities.

I have pointed out that one set of practices-or-abilities can be elaborated into another by a process of *training*, rather than algorithmically, and that the practices-or-abilities that implement algorithmic elaboration are neither necessary nor sufficient for this sort of practical elaboration. Besides this negative characterization, what can we say positively about what training is? Most generally, I think of training as a course of *experience*, in Hegel's and Dewey's sense (processual, developmental *Erfahrung* rather than episodic, self-intimating *Erlebnis*) of a feedback loop of perception, responsive performance, and perception of the results of the performance. When we think about the practices-or-abilities that *implement* elaboration-by-training, we can think about them on the side both of the trainer and of the trainee (though both learning—training without a trainer—and self-training, which is not the same thing, are also important species). A course of training implements a *pedagogical* elaboration of one set of abilities into another. We can think of it very abstractly as having as its basic unit a stimulus (perhaps provided by the trainer), a response on the part of the trainee, a response by the trainer to that response, and a response to that response by the trainee that involves altering his dispositions to respond to future stimuli. A constellation of such units constitutes a *course* of *training*.

6 Algorithmic pedagogical decomposition and pedagogical politics

I am suggesting that what, in a course of training, is most analogous to algorithmic elaboration of abilities is *pedagogical* elaboration in the form of a training regimen. In rare but important cases in early education, we have *completely solved* the problem of how to pedagogically elaborate one set of abilities into another. What it means to have a *solved pedagogical problem* for a population with respect to an output practice-or-ability is to have an *empirically sufficient conditional branched training regimen* for it. This is something that as a matter of contingent fact can take any novice from the population who has mastered the relevant range of primitive practical capacities, and by an algorithmically specifiable Test-Operate-Test-Exit (TOTE) cycle of responses to her responses *in fact* (though without the guarantee of any principle), get her to catch on to the target ability. For us, training pupils who can already *count* to be able to *add* is essentially a

solved pedagogical problem in this sense. That is, starting with pupils of widely varying abilities and prior experiences, who share only the prior ability to count, there is a flowchart of differentially elicited instructions, tests, and exercises that will lead all of them to the target skill of being able correctly to add pairs of arbitrary multi-digit numbers. A common initial lesson or exercise is followed by a diagnostic test. The results of that test then determine, for each pupil, which of an array of possible second lessons or exercises is appropriate, followed by further tests whose results are interpreted as differentially calling for different exercises, and so on. This flowchart determines a TOTE cycle of training that incorporates a *pedagogical* (as opposed to an executive) *algorithm*.

I am told by those who know about these things that teaching *multiplication* to pupils who can add and subtract is also, in this sense, a completely solved pedagogical problem, but that in spite of massive investigative efforts to date, *subtraction* remains an essentially *un*solved pedagogical problem, and *division*, in the form of mastery of fractions, a tantalizing, so far intractable pedagogical *mystery*. In the absence of a complete practical pedagogical algorithm, those charged with eliciting and developing such skills must fall back on rougher heuristics and the sort of practical know-how gleaned from many years of trial-and-error training of a wide variety of candidates.

Incompletely solved pedagogical problems—not just in specialized cases in elementary education, but at all levels and across the board—raise a broad issue of institutional *politics* that seems to me to penetrate deeply into our understanding of, and attitudes towards, the society as a whole. Augustine marveled at (and rode three days on a mule to test) the rumored ability of a monk to gather the sense of a text without pronouncing aloud the words on the page and then listening to them. One of Samuel Pepys's distinctive qualifications for his position as Secretary of the Admiralty was his mastery of the arithmetic required for double-entry bookkeeping. Today we take it for granted that we can train almost everyone to read silently and to add up long columns of figures. But for abilities for which the pedagogical problem has not been completely solved, where we do not yet have an algorithmic decomposition of the practical training process, candidates who exhibit all the relevant primitive abilities are *de facto sorted* by the training regimens we do have, not only by the number of iterations of the TOTE loop it takes for them to acquire the target ability, but also by whether they can be brought to that point at all. Matter-of-factual

PP-necessity relations among practices-or-abilities require that the outputs of some training regimens serve as the inputs to others—that some of the abilities treated as primitive (as practically a priori) by the one are achievable only as the target abilities (the practical a posteriori) of others. It follows that the effects of failing to acquire one ability—falling into the missing, incompletely mapped portion of an ideally complete pedagogical solution of which some actual training regimen is a mere fragment—will be strongly cumulative within a sequence of courses of learning-and-training.

The broadly political issue I want to point to concerns how, in the context of these very general considerations, we should think about one element of just treatment of individuals by institutions. We might, as a demand of justice, or simply as a counsel of social engineering, want some kinds of rewards to be proportioned to productive achievements, according to some definition of the latter. Among the crucial necessary conditions of any such achievement is the possession of certain skills or abilities. It seems that there are two basic attitudes (defining a spectrum between them) that one might have toward any target ability for which we do *not* have a pedagogical algorithm codifying a complete solution to the training problem.

One attitude is that it is just a brute empirical fact that people not only have different abilities, but are in important respects more or less able. With respect to any sort of target ability, some are more trainable, better learners, than others. What is being assessed here is the practical-elaborative abilities that *implement* the PP-sufficiency of some set of primitive abilities for the target ability, in the context of the course of experience yielded by a training regimen. The training regimen not only inculcates or elicits the skill that is its target, but along the way sorts candidates into those who can and those who cannot learn or be trained in it, as well as into those who learn it faster or more easily—measured by how long or how many steps it takes to get them through the pedagogical flowchart to the exit of that practical labyrinth. On this view, it is *compatible* with just dealing, and perhaps even *constitutive* of a dimension of justice, for an institution to factor this sort of second-order ability into its reward structure.

The view that forms the opposite pole of the dimension I am pointing to focuses on the relativity of the hierarchical sorting of candidates into more or less trainable to the training regimens that happen to be available. Different regimens might produce quite different rankings. If that is so, and the fact that we have actually implemented one set of training procedures

rather than another is quite contingent, conditioned by adventitious, in principle parochial, features of the actual history of the training institutions, the experience and skills the available trainers happen to have, and so on, then the inferences from the actual outcomes of training either to the attribution of some kind of *general* second-order ability or to the *justice* of rewarding the *particular* sort of second-order ability that really *is* evidenced thereby—just being more trainable, or more easily trainable, by the methods we happen to have in place and apply—are undercut. Our failure to provide a more comprehensive set of training alternatives, to have filled in the pedagogical flowchart more fully, ultimately, to have completely solved the relevant training problem, should be held responsible for the training outcome, rather than supposing that sub-optimal outcomes reveal evaluatively significant deficiencies on the part of the trainee. At the limit, this attitude consists in a *cognitive* commitment to the effect that there is, in principle, a complete pedagogical algorithmic solution for every target skill or ability to the possession of which it is just to apportion rewards, and a *practical* commitment to find and implement those solutions. It is an extreme, indeed utopian, pedagogical egalitarianism.

Taken to the limit, the pedagogical egalitarian view may seem to rest on a literally unbelievable premise: that whatever *some* human can (learn or be trained to) do, *any* human can (learn or be trained to) do. And the evaluative component implicit in the cognitive commitment as I stated may seem no more plausible: that there is something wrong with rewarding any ability of which that claim is *not* true. A more defensible version of pedagogical egalitarianism results if the latter commitment is softened so as to claim merely that special arguments must be given in each case for the valorization of differences in ability for which we have found no complete pedagogical solution. The first element of the cognitive component can correspondingly be interpreted in terms of the practical commitment, so as to claim that we have no right to assume, for any given skill or ability for which we have as yet no complete pedagogical solution, that that is because there *is*, in principle, no such solution.

The empirical-hierarchical attitude is *conservative* in treating extant institutionalized training regimens as given and fixed, and the utopian pedagogical egalitarian attitude is *progressive* in its commitment to transform them. As political attitudes, they articulate one dimension of the nature/nurture aspect of traditional right/left alignments. Between them

lies a whole array of more nuanced principles for assigning reciprocal, co-ordinate responsibility to training or trainers, on the one hand, and trainees on the other. We need not simply choose between the strategies of holding actual training regimens fixed and hierarchically sorting humans with respect to them, on the one hand, and holding the actual practical-elaborative abilities of humans fixed and sorting training regimens with respect to them, on the other.

But my purpose in gesturing at this issue of pedagogical politics here has not been to recommend one or another way of approaching it. Assessing the plausibility of the broadened, practical version of the thesis of artificial intelligence led to the notion of practical PP-sufficiency by training. My aim in this final section has been to lay alongside the postulate of universal practical *executive* algorithmic decomposability of discursive abilities, characteristic of AI-functionalism, the postulate of universal practical *pedagogical* algorithmic decomposability of discursive abilities characteristic of utopian pedagogical egalitarianism, and to point to an issue of considerable philosophical, cultural, and political significance that it raises. As a result, the argument of the lecture as a whole has described a narrative arc taking us from Turing, through Wittgenstein, to Dewey.

My last three lectures will address *modal* vocabulary, *normative* vocabulary, and the pragmatically mediated semantic relations they stand in to ordinary *objective, empirical*, and *naturalistic* vocabularies, and to each other. I will argue that both the *deontic* vocabulary of conceptual norms and the *alethic* vocabulary of laws and possibilities can be elaborated from and are explicative of features necessarily exhibited by any autonomous discursive practice. Thinking about the pragmatically mediated semantic relations they stand in to each other turns out to provide a new way of understanding the *subjective* and *objective* poles of the intentional nexus of knowers-and-agents with their world. Along the way, I will show how *normative* vocabulary can serve as an expressively bootstrapping pragmatic metavocabulary for *modal* vocabulary, and, in the fifth lecture, how that fact makes possible a new sort of *formal semantics* for logical and modal vocabulary, as well as for ordinary empirical descriptive vocabulary.

4

Modality and Normativity: From Hume and Quine to Kant and Sellars

1 The modal revolution

The status and respectability of alethic *modality* was always a point of contention and divergence between naturalism and empiricism.[1] It poses no problems in principle for *naturalism*, since modal vocabulary is an integral part of all the candidate naturalistic base vocabularies. Fundamental physics is, above all, a language of *laws*; the special sciences distinguish between true and false *counterfactual* claims; and ordinary empirical talk is richly *dispositional*. By contrast, modality has been a stumbling block for the *empiricist* tradition ever since Hume forcefully formulated his epistemological, and ultimately semantic, objections to the concepts of <u>law</u> and <u>necessary connection</u>.

Those traditional reservations about the intelligibility of modal notions were underscored, reinforced, and confirmed for twentieth-century versions of empiricism, which had been distinguished, strengthened, and made more precise by the addition of the semantic logicist model of the conceptual articulation of empirical content. Extensional, first-order quantificational languages could express *regularities* and *generalizations* with hitherto undreamed of power and precision. But for philosophers from Russell through Carnap to Quine, that just made it all the more urgent to explain, or explain away, the *lawlikeness* or counterfactual-supporting *necessity* distinctive of at least *some* of those generalizations, which demonstrably

---

[1] This tension was a principal source of conflict within the Vienna Circle, dividing Neurath and Schlick, for instance, with Carnap trying to mediate.

extended beyond what can be captured by the expressive resources of that logical vocabulary.[2]

This confluence of traditional empiricist with logicist difficulties concerning the content expressed by modal vocabulary had the result that for roughly the first two-thirds of the twentieth century, Anglophone philosophy regarded alethic modal vocabulary with extreme suspicion, if not outright hostility. It ranked, with normative vocabulary, as among the most mysterious and philosophically puzzling forms of discourse, the source of central standing and outstanding philosophical problems, as a prime candidate for the analytic project of semantic clarification in favored terms or, failing that, principled elimination from perspicuous discourse, as Quine famously recommended.

But philosophical attitudes towards modality underwent a remarkable, in many ways unprecedentedly radical, transformation during the twentieth century. For, starting in the second half of the century and accelerating through the last third, modal vocabulary became the analytic semanticist's best friend, and an essential part of the contemporary philosopher's metaconceptual toolkit. I think it is worthwhile reminding ourselves just how surprised and astonished philosophers who lived and moved and had their being in the earlier milieu would have been to discover that, by the end of their century, when questions were raised about the semantics of some vocabulary—for instance, normative, intentional, or even semantic vocabulary itself—not only the dominant strategy, but the very first recourse would be to appeal to *modal* notions such as dispositions, counterfactual dependencies, and nomological relations to explain the questionable conceptual contents. Just how—they would want to know—did what seemed most urgently in need of philosophical explanation and defense suddenly become transformed so as to be unproblematically available to explain other puzzling phenomena? Surely such a major transformation of *explanandum* into *explanans* could not be the result merely of a change of fashion, the onset of amnesia, or the accumulation of fatigue? But if not, what secret did we find out, what new understanding did we achieve, to *justify* this change of philosophical attitude and practice?

---

[2] We now know, thanks to Danielle Macbeth's *Frege's Logic*, that Frege's own *Begriffsschift* notation did not share the expressive impoverishment with respect to modality exhibited by the extensional first-order logic that Russell, and following him, everyone else, drew from it.

Two answers to this question lie ready to hand. First, there was a formal-semantic revolution in modal logic. And second, the Anglophone tradition more or less gave up empiricism in favor of naturalism. I think both those explanations are right, as far as they go, both as a matter of historical fact and in the order of justification. But it is important to understand exactly *which* questions those developments *did* offer responsive answers to, and to which they did *not*.

As to the first point, I think there is a widespread tendency to think that, to paraphrase Alexander Pope:

> Modality and Nature's laws lay hid in night,
> God said: "Let Kripke be!" and all was light.

But that cannot be right. Kripke's provision of a complete extensional semantic metavocabulary for intensional modal logical vocabulary—and its powerful development, by others such as Montague, Scott, Kaplan, Lewis, and Stalnaker, into a general intensional semantics for non-logical vocabulary—*is* an adequate response to worries stemming from the *extensional* character of the *logical vocabulary* in which semantics had been conducted. That is, it addresses the difficulties on the *semantic logicist* side of the classical project of analysis that stem from the expressive impoverishment of first-order logical vocabulary. But these formal developments do *not* provide an adequate response to residual *empiricist* worries about the intelligibility of modal concepts. For the extensionality of the semantic metalanguage for modality is bought at the price of making free use of modal primitives: most centrally, the notion of a <u>*possible* world</u> (as well as that of <u>accessibility</u> relations among such *possibilia*). As Quine emphasized, the modal vocabulary whose use is essential to this semantic approach evidently falls within the circle of terms and concepts to which empiricist suspicions and questions apply. That is, even putting *ontological* issues aside, whether possible worlds are thought of as abstract objects, as concrete particulars spatio-temporally unconnected to our universe, or as *sui generis possibilia*, the *epistemological* question of how we are to understand the possibility of our *knowing* anything about such items (and their accessibility relations), and the question how, if the possibility of such *cognitive* contact is mysterious, the idea of our having the *semantic* contact necessary so much as to *talk* or *think* about them can be made intelligible, are wholly

untouched by this formal apparatus, and remain every bit as pressing as before.

2 The modal Kant–Sellars thesis

How urgent those questions are depends on whether we have grounds to accept criticisms of the empiricist program that undermine the basis for its relegation of modal vocabulary to a suspect, second-class status. I think that the best justification for our new comfort with modal idioms is indeed to be found in the principled rejection of some of the crucial presuppositions of the empiricist critique of the credentials of modal concepts. We can now see that the operative core of both Quine's and Sellars's arguments against empiricism consists in objections to its underlying *semantic atomism*.[3] Arguing that meaning must at least determine inferential role, and noticing that what follows from or is evidence for or against a claim depends on what other claims are available as auxiliary hypotheses or collateral premises, Quine concludes that the smallest unit of meaning is not a sentence, even in the case of observation sentences, but what he calls a 'theory': the whole constellation of all sentences held true, articulated by their inferential relations both to one another and to sentences not held true. Sellars argues that even observational beliefs acquired non-inferentially through perception can be understood as *conceptually* contentful—and hence potentially cognitively significant—only in virtue of their inferential relations to other possible beliefs. He concludes that non-inferential reports, no matter what their subject matter, cannot constitute an autonomous discursive practice: a language-game one could play though one played no other.

It is clear, I take it, how these anti-atomist arguments bear against empiricist foundationalism: the layer-cake picture of a semantically autonomous base of perceptual experience or reports thereof, on which is erected a semantically optional superstructure, in effect, of theories inferentially based on those observations. And insofar as empiricist worries about the status of laws, necessary connections, dispositions, and counterfactual possibilities

---

[3] In their classic papers of the 1950s, ''Two Dogmas of Empiricism'' (1953) and ''Empiricism and the Philosophy of Mind'' (1956).

are predicated on the difficulty of justifying the inferences that would add them to the supposedly semantically autonomous base of non-modal reports of actual experiences, Quine's and Sellars's assault on the layer-cake picture, if successful, undercuts those worries by removing the motivation for their ultimately unmeetable constraints on an account of what modal vocabulary expresses. Thought of this way, though, criticism of the semantic presuppositions of the empiricist project does not bear any more directly on its treatment of *modal* vocabulary than it does on its treatment of any other potentially puzzling candidate for empiricist explication: *theoretical* (that is, non-observational, exclusively inferentially applicable) vocabulary, *normative* vocabulary, *probabilistic* vocabulary, and so on.

But there is another, much more intimate and immediate *positive* connection between arguments against semantic atomism and our understanding of what is expressed by the use of modal vocabulary. And it is here that I think we can find the best justification for our current relaxed attitude toward, and even enthusiastic embrace of, modal idioms as suitable tools for serious analytic semantic work. The underlying idea is what I will call the "Kant-Sellars thesis about modality." Hume found that even his best understanding of actual observable empirical *facts* did not yield an understanding of *rules* relating or otherwise governing them. Those facts did not settle which of the things that *actually* happened *had* to happen (given others), that is, were (at least conditionally) *necessary*, and which of the things that did *not* happen nonetheless were *possible* (not ruled out by laws concerning what did happen). Though initially couched as an *epistemological* question about how one could *know* what rules or laws were in play, Hume's worries run deeper, raising the *semantic* question of what it could so much as *mean* to say that the facts are governed or related by rules or laws. Hume (and, following him, Quine) took it that epistemologically and semantically fastidious philosophers faced a stark choice: either show how to explain modality in non-modal terms or learn to live without it. But that challenge is predicated on the idea of an independently and antecedently intelligible stratum of empirical discourse that is purely descriptive and involves no modal commitments, as a semantically autonomous background and model with which the credentials of modal discourse can then be invidiously compared. One of Kant's most basic ideas, revived by Sellars, is that this idea is mistaken. **The ability to use ordinary empirical descriptive terms such as 'green', 'rigid', and 'mass' already presupposes grasp**

of the kinds of properties and relations made explicit by modal vocabulary. Sellars summed up the claim admirably in the title of one of his early papers: "Concepts as Involving Laws, and Inconceivable Without Them."[4]

Kant was struck by the fact that the essence of the Newtonian concept of <u>mass</u> is of something that, by law, *force* is both necessary and sufficient to *accelerate*. And he saw that all empirical concepts are like their refined descendants in the mathematized natural sciences in this respect: their application implicitly involves counterfactual-supporting dispositional commitments to what *would* happen *if*. Kant's claim, put in more contemporary terms, is that an integral part of what one is committed to in applying any determinate concept in empirical circumstances is drawing a distinction between counterfactual differences in circumstances that *would* and those that *would not* affect the truth of the judgment one is making. One has not grasped the concept <u>cat</u> unless one knows that it would still be possible for the cat to be on the mat if the lighting had been slightly different, but not if all life on earth had been extinguished by an asteroid-strike.[5]

In an autobiographical sketch, Sellars dates his break with traditional empiricism to his Oxford days in the 1930s. It was, he says, prompted by concern with the sort of content that ought to be associated with logical, causal, and deontological modalities. Already at that point he had the idea that

---

[4] First published in 1948; reprinted at pp. 87–124 in J. Sicha (ed.), *Pure Pragmatics and Possible Worlds: The Early Essays of Wilfrid Sellars* (Ridgeview Publishing Co., 1980), hereafter *PPPW*. This slogan is a good place to start in thinking about Kant's point, but in fact Sellars's own view is subtly but importantly different from Kant's. For Sellars, the laws determining the truth of counterfactuals involving the application of a concept are part of the content of the concept. For Kant, modal concepts make explicit not something implicit in the *content* of determinate concepts, but something implicit in their *empirical use*, in *applying* them to make empirical *judgments*. That is why the pure concepts of the understanding—what he calls 'categories', such as <u>possibility</u> and <u>necessity</u>—both are to be understood in terms of the forms of judgment (the table of categories derives from the table of judgments) and express synthetic, rather than analytic, necessities. From Kant's point of view, a better slogan than Sellars's would be "The *Use* of Concepts in Empirical *Judgments* as Involving Laws and Inconceivable Without Them."

[5] It is this observation, unwittingly underscored by Hume (for Kant, the Moses who brought us to within sight of the Promised Land he himself was destined not to enter), that motivates Kant to wheel in his heavy transcendental machinery. For Kant sought to explain the modal commitments implicit in the application of ordinary empirical concepts by placing the modal concepts of <u>law</u> and <u>necessity</u> in the newly postulated realm of *pure* concepts or categories, which must be graspable a priori precisely in the sense that their applicability is presupposed by the applicability of *any* empirical concepts. The concept of vocabularies that are "universally LX," introduced below, is a successor notion along at least one important dimension.

what was needed was a functional theory of concepts which would make their role in reasoning, rather than supposed origin in experience, their primary feature.[6]

Somewhat more specifically, he sees modal locutions as tools used in the enterprise of

making explicit the rules we have adopted for thought and action. … I shall be interpreting our judgments to the effect that A causally necessitates B as the expression of a rule governing our use of the terms 'A' and 'B'.[7]

In fact, following Ryle,[8] he takes modal expressions to function as *inference licenses*, expressing our commitment to the goodness of counterfactually robust inferences from necessitating to necessitated conditions. If and insofar as it could be established that their involvement in such counterfactually robust inferences is essential to the *contents* of ordinary empirical concepts, then what is made explicit by modal vocabulary is implicit in the use of any such concepts. That is the claim I am calling the "Kant-Sellars thesis." On this view, modal vocabulary does not just add to the use of ordinary empirical observational vocabulary a range of expressive power that is *extraneous*—as though one were adding, say, *culinary* to *nautical* vocabulary. Rather, the expressive job distinctive of modal vocabulary is to articulate just the kind of essential semantic connections among empirical concepts that Sellars (and Quine) point to, and whose existence semantic atomism is principally concerned to deny.

As I would like to formulate it, the Kant-Sellars thesis begins with the claim that in using ordinary empirical vocabulary, one already knows how to do everything one needs to know how to do in order to introduce and deploy modal vocabulary. If that is right, then one cannot be in the position the atomist (for instance, empiricist) critic of modality professes to find himself in: having fully understood and mastered the use of *non*-modal vocabulary, but having thereby afforded himself no grip on the use of *modal* vocabulary, and no access to what it expresses. The Humean-Quinean predicament is accordingly diagnosed as resulting from a failure properly to understand the relation between modal vocabulary and what one must *do* in order to *deploy* non-modal, empirical, descriptive vocabulary.

---

[6] In H. N. Castañeda (ed.), *Action, Knowledge, and Reality* (Bobbs-Merrill, 1975), 285.

[7] Sellars, "Language, Rules, and Behavior" (1949), reprinted in *PPPW*, 136, fn. 2.

[8] Gilbert Ryle, " 'If', 'So', and 'Because'," in Max Black (ed.), *Philosophical Analysis* (Prentice Hall, 1950), 302–18.

The thought that the expressive role characteristic of alethic modal vocabulary is to make explicit semantic or conceptual connections and commitments that are already implicit in the use of ordinary (apparently) non-modal empirical vocabulary faces at the outset at least two sorts of potentially weighty objection. First, did not Kripke's semantic investigations of modally rigid designators reveal the sort of necessity they articulate as being *metaphysical*, specifically by contrast to the sort of *conceptual* necessity that Quine, for instance, had worried about and rejected? And second, to talk about what is necessary and possible is not to *say* anything about rules for using linguistic expressions, or about what anyone is committed to, since the objective modal claims in question could have been true even if there had never been language users, linguistic expressions, rules, or commitments.

As to the first objection, the philosophical phase of the modal revolution (developing the earlier logical and semantic phases of that revolution) that Kripke precipitated in ''Naming and Necessity''[9] did indeed use the semantic phenomenon of the modal rigidity of some non-descriptive vocabulary to articulate a kind of necessity that is knowable only a posteriori. The conclusion that such necessity should not be understood as *conceptual* necessity follows only if one either identifies conceptual content with *descriptive* content (by contrast to the causally-historically acquired content of proper names and demonstratives) or takes it (as Quine, following the tradition, had) that conceptual connections must be knowable a priori by those who have mastered those concepts. But both of these are optional commitments, which can and should be rejected by anyone trying to follow out the Kant-Sellars line of thought about modality. McDowell has argued, to my mind convincingly, that the content expressed by demonstrative vocabulary should be understood as thoroughly conceptual (and that Frege already took it to be so).[10] And in *Making It Explicit*, I articulate a broadly inferential notion of the conceptual that incorporates the indirectly inferential roles of substitution and anaphora—including the anaphoric phenomenon that is modal rigidity.[11]

---

[9] In Donald Davidson and Gilbert Harman (eds.), *Semantics of Natural Language.* Dordrecht: Reidel, 1972.

[10] ''*De Re* Senses,'' in *Meaning, Knowledge, and Reality* (Harvard University Press, 2001).

[11] Robert Brandom, *Making It Explicit* (Harvard University Press, 1994), chapters 6, 7 (especially sections III and IV), and 8 (section V).

On the other point, Sellars's forthright response to Quine's pragmatic challenge in "Two Dogmas of Empiricism"—to say what it is about the *use* of expressions that distinguishes inferences underwritten by necessary conceptual relations from those underwritten by contingent matter-of-factual ones—is to identify the concept-articulating inferences as those that are counterfactually robust.[12] He cheerfully embraces the consequence that to discover what is contained in the concept copper one needs empirically to investigate the laws of nature. (This is a kind of semantic 'externalism' that does not need to take on the dangerous and difficult task of making sense of a notion of the 'internal' with which to contrast.) The issue about conceptual necessities here is not an empirical one: who is right about the conceptual? The Kant-Sellars thesis about modality requires deploying a concept of the conceptual that differs in important ways from the traditional one. As long as such a notion can be intelligibly developed and consistently applied, those differences need only be kept firmly in mind, not counted as fatal flaws.

The response to the second objection (that saying what is necessary or possible is not saying anything about how anyone talks) must be to be clearer about the sort of pragmatically mediated semantic relation the Kant-Sellars thesis takes modal vocabulary to stand in to ordinary, non-modal descriptive vocabulary. The large claim in the vicinity—one that will occupy me not only in this lecture but beyond—is, as Sellars puts it, that "the language of modality is…a 'transposed' language of norms."[13] I do not think that Sellars himself ever manages to say clearly just what sort of 'transposition' he has in mind. He appeals to a distinction between what is *said* by the use of some vocabulary, and what is *conveyed* by its use. While admitting that talk of what is necessary does not *say* anything about what language users ought or ought not to do, he nonetheless insists that it "conveys the same information" as "rules to the effect that we may do thus and so, and ought not do this and that, in the way of manipulating expressions in a language."[14] His (only somewhat helpful) example is that when I say "The sky is clear," I have both said something about the weather and conveyed something about my beliefs. The point, I take it, is to distinguish what follows *semantically* from the content of what I have

---

[12] Sellars, "Is There a Synthetic *A Priori*?" *Philosophical Studies*, 20 (1953), 121–38.

[13] Sellars, "Inference and Meaning" (1953), reprinted in Kevin Scharp and Robert Brandom (eds.), *In the Space of Reasons* (Harvard University Press, 2007), 21.

[14] Ibid.

said from what follows *pragmatically* from my *saying* of it. (Embedding the claims as the antecedents of conditionals will distinguish these two sorts of consequences. "If the sky is clear, then it will not rain," expresses a good inference, whereas "If the sky is clear, then Brandom believes that the sky is clear," does not. For only the semantic content, and not the pragmatic force of the utterance, survives such embedding.)

We are in a position to be a little clearer about what Sellars is after with his dark notion of what an utterance 'conveys'. The view is that what I am *doing* when I say that it is causally necessary that if this piece of copper is heated to 1084 °C, it will melt, is endorsing a certain kind of inference. I am not *saying that* that inference is good; the facts about copper would be as they are even if there were no inferrers or inferrings. When Sellars says "the language of modality is … a 'transposed' language of norms," he is saying in my terms that normative vocabulary codifying rules of inference is a *pragmatic metavocabulary* for modal vocabulary. His 'transposition' is just this pragmatically mediated semantic relation between deontic normative and alethic modal vocabulary. The corresponding meaning-use diagram (MUD) is shown in Figure 4.1.

In Lecture 5, I will show in detail how this thought can be exploited to develop a new sort of formal semantics, which yields new insights into the conceptual contents expressed by ordinary empirical-descriptive vocabulary, as well as logical and modal vocabularies.

This claim is merely part of the background of what I have been calling the "Kant-Sellars thesis" about modality, however. That thesis comprises two claims:
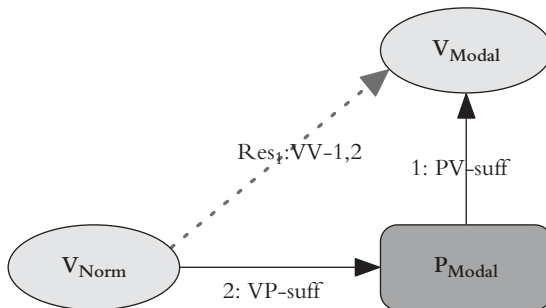
Figure 4.1 "The language of modalities is a 'transposed' language of norms"

1. In using ordinary empirical vocabulary, one already knows how to do everything one needs to know how to do in order to introduce and deploy *modal* vocabulary.
2. The expressive role characteristic of alethic modal vocabulary is to *make explicit* semantic, conceptual connections and commitments that are already *implicit* in the use of ordinary empirical vocabulary.

The first says that some practices that are PV-necessary for the use of any empirical vocabulary are PP-sufficient for practices that are PV-sufficient to deploy modal vocabulary. The second says that that modal vocabulary then makes explicit those aspects of practices-or-abilities that are implicit in the use of any empirical vocabulary. In fact, these are ways of saying that modal vocabulary stands to ordinary empirical vocabulary in the complex, pragmatically mediated semantic relation I have already identified (in Lecture 2) as elaborating-explicating: the meaning-use relation called 'LX' for short. The corresponding MUD is shown in Figure 4.2.

Combining these claims yields a MUD asserting relations among modal, normative, and empirical vocabularies (Figure 4.3).

3 Counterfactual robustness and the updating argument

So far, I have only expounded, explicated, and mentioned some of the consequences of the Kant-Sellars thesis about modal vocabulary, but not
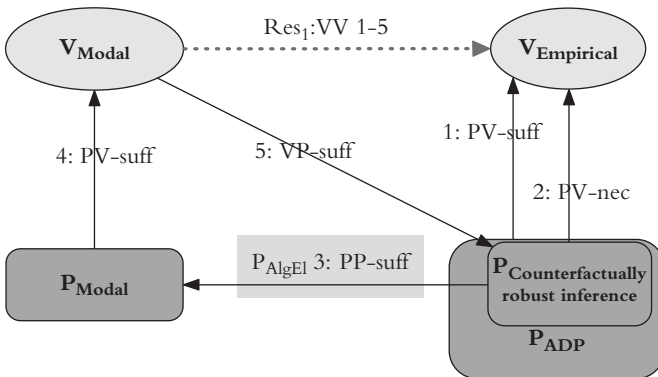
Figure 4.2 The Kant-Sellars thesis: modal vocabulary is elaborated-explicating (LX)
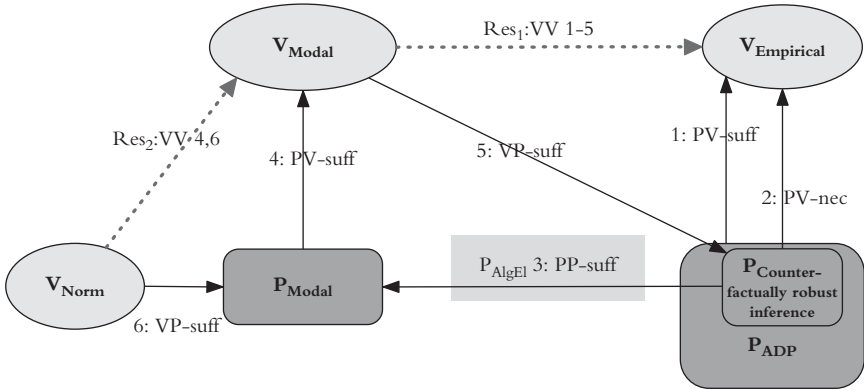
Figure 4.3 Modal, normative, and empirical vocabulary

sought to *argue* for it. What reason is there to think that it is *true*? The analysis of the Kant-Sellars thesis as asserting a complex pragmatically mediated semantic relation between vocabularies that is the resultant of a definite constellation of basic meaning-use relations, as presented in the MUD, tells us exactly what shape such an argument must have. For it tells us just which *basic* meaning-use relations must be established in order to show that the *resultant* one obtains. The key element in this case will be finding some set of practices that can be argued to be at once contained in or exhibited by every autonomous discursive practice, and PP-sufficient for practices PV-sufficient for deploying explicitly modal vocabulary, which is VP-sufficient to specify the original PV-necessary practices-or-abilities. As the labels on the MUDs indicate, for the argument I will mount (picking up on the theme on which my argument last time turned), those practices are *counterfactually robust inferential* practices-or-abilities—more specifically, the practical capacity to associate with materially good inferences *ranges of counterfactual robustness*. If it can be established that deploying any ordinary empirical vocabulary presupposes these practices-or-abilities, and that they in turn suffice to introduce explicit modally qualified conditionals that permit the expression of those practical discriminations, then the universal LX-character of modal vocabulary relative to ordinary empirical vocabulary will have been demonstrated.

I have already claimed that any autonomous discursive practice (ADP) must include practices-or-abilities of distinguishing some inferences as materially good from others that are not. For some bit of vocabulary to

function as a propositionally contentful declarative sentence is for it to be available to serve as the premise and conclusion of such material inferences. Further, it is the expressive job generically characteristic of *conditional* vocabulary to *codify* endorsements of material inferences: to make them explicit in the form of declarative sentences that can themselves serve as the premises and conclusions of inferences. The philosopher most responsible for getting us to think about conditionals in this way is Gilbert Ryle. In his classic essay " 'If', 'So', and 'Because'," in which he introduces the idea of hypothetical statements as inference tickets or licenses, he also points out an intimate connection between them and *modal* claims. He says:

We have another familiar way of wording hypothetical statements. Although the standard textbooks discuss "modal propositions" in a different chapter from that in which they discuss hypotheticals, the differences between modal and hypothetical statements are in fact purely stylistic. There is only one colloquial way of correctly negating the superstitious hypothetical statement "If a person walks under a ladder, he comes to grief before the day is out," namely, by saying "No, a person may (might, or could) walk under a ladder and not come to grief." And the only colloquial way of putting a question to which an "if-then" statement is the required affirmative answer is to ask, for example, "Can an Oxford Vice-Chancellor not be (or need he be) a Head of College?" … [W]e always can reword an "if-then" statement as a statement of the pattern "It cannot be Monday today and not be Tuesday tomorrow."[15]

I think he is right that "It is possible that (p and *not-q*)" is incompatible with "if p then q" when the latter is used to codify an ordinary material inference such as the inference from a banana's being yellow to its being ripe. Endorsing a material inference does involve a commitment of the sort made explicit by the use of modal vocabulary, about what is and is not possible, and what is at least conditionally necessary.

For this reason, the fact that we cannot intelligibly describe someone as deploying a concept unless he makes some distinction between materially good and bad inferences involving it has the consequence that we also cannot understand the practitioner as deploying the concept unless he treats the material inferences he takes to be good as having a certain *range of counterfactual robustness*, that is, as remaining good under various merely

---

[15] Gilbert Ryle, " 'If', 'So', and 'Because'," 313.

hypothetical circumstances. One grasps the claim "the lioness is hungry" only insofar as one takes it to have various consequences (which *would* be true if it *were* true) and to rule out some others (which *would not* be true if it *were* true). And it is not intelligible that one should endorse as materially good an inference involving it, such as the inference from "the lioness is hungry" to "nearby prey animals visible to and accessible by the lioness are in danger of being eaten," but be disposed to make no distinction at all between collateral premises that would, and those that would not, if true infirm the inference. One must make *some* distinction such as that the inference would still go through if the lioness were standing two inches to the east of her actual position, the day happened to be a Tuesday, or a small tree ten miles away cast its shadow over a beetle, but not if she were shot with a tranquilizing dart, the temperature instantly plummeted 300 degrees, or a plane crashed, crushing her. The claim is not that one could not fail to assess some, or even all, of *these particular* counterfactuals correctly and still count as grasping the claim that is their premise, but that one could not so qualify if one made *no* such distinctions.

It may initially be tempting to think that the inferences that are counterfactually robust are all and only those underwritten by *laws*. Thus inferences underwritten by the law that all samples of copper melt at 1083.4 °C are counterfactually robust: if this coin (which in fact is silver) *were* made of copper, it *would* melt at 1083.4 °C. Whereas inferences underwritten by the accidental regularity that all the coins in my pocket are copper are not counterfactually robust: if I *were* to put this coin (which in fact is silver) in my pocket, it would *not* be copper. There are indeed real and significant differences between these cases, but I think it is important not to think of them in terms of the difference between inferences that *are* counterfactually robust and inferences that are *not*. The difference is rather one of the character of the particular *ranges* of counterfactual robustness. For the accidental generalization that all the coins in my pocket are copper *does* underwrite counterfactuals such as: "If I *were* to choose a coin at random from my pocket, it *would* be copper." In fact, *every* claim, whether contingent or not, supports *some* counterfactual inferences, and if one grasped *none* of them one would not qualify as understanding those claims.

I think these considerations suffice to establish that autonomous discursive practices *essentially*, and not just *accidentally*, involve the association of ranges of counterfactual robustness with at least some material inferences. If, as

Ryle claims, and as is in any case plausible, *modal* vocabulary specifying what is, at least, conditionally possible and necessary can then be introduced to make explicit those commitments to the, at least, limited counterfactual goodness of material inferences, then we have what is needed for the modal Kant-Sellars thesis. But I think that if we dig deeper, we can learn more. So, rather than leaving things at this point, I want to consider a more detailed line of argument for this, the most potentially controversial element of the complex meaning-use relation that thesis asserts.

For the first premise, I take it to be clear that every autonomous discursive practice must have some vocabulary that can be used *observationally*, in reliably differentially elicited non-inferential reports. This is the core of what I have been referring to as "ordinary empirical vocabulary." Second, I have already argued that those who engage in any discursive practices must distinguish in practice between materially good and materially bad inferences—where calling them 'material' just means that the presence of some non-logical vocabulary is essential to the classification. Recall that this is not to claim that they must have a view about the goodness or badness of every possible candidate material inference; there can be some about which they have no view. And it is not to claim that they always are correct about the goodness of the inferences toward which they do have attitudes. But to count as deploying any vocabulary at all, one must treat some inferences involving it as good and others as bad. Otherwise, one's utterances are wholly devoid of conceptual content; whatever pragmatic significance they may have, it cannot be thought of as *discursive* significance. Even tokenings that are non-inferentially elicited by environing stimuli—that is, the applications of observational vocabulary mentioned in the first premise—must have inferential *consequences*, if they are not to be cognitively idle.

The third claim is that material inference is in general *non-monotonic*. That is, the inference from p to q may be materially good, even though the inference from $p \, \& \, r$ to q is not. Monotonicity of inference is, of course, a familiar feature of inferences within a formal *logical* system, and in mathematical reasoning; and that feature is arguably inherited by fundamental physics. But in the special sciences inferences are almost always *defeasible*, by collateral circumstances that thereby count as 'special'. Each stage in a physician's differential diagnosis is like this: the inference from test result, physical finding, or symptom is surrounded by a nimbus

of usually unspoken 'unless'es. And no one supposes that such probative reasoning can always be turned into dispositive reasoning by making an explicit, exhaustive list of the potential defeasors. Certainly, reasoning in everyday life does not generally admit such completions. If I strike this dry, well-made match, it will light—unless it is done inside a strong magnetic field. But it still will light if, in addition, it is struck inside a Faraday cage—unless there is not enough oxygen. And so on. There need be no definite totality of possible defeasors, specifiable in advance. Even where we have some idea how to enumerate them, unless those provisos are generally left implicit, actually stating the premises so as to draw inferences from them monotonically is impossibly cumbersome in practice.

At this point, one is liable to think of *ceteris paribus* clauses. The careful way to formulate the ordinary inference just mentioned is that if I strike this dry, well-made match, *ceteris paribus*, or other things being equal, it will light. I think that is indeed exactly what we ought to say, and the point I want to make can be made by saying that what such *ceteris paribus* clauses mark is an unavoidable feature of ordinary material inferences. But it is critical to understand what such clauses *do* and do *not* do. They are *not* devices for the wholesale stipulation of the denial of all of the potential defeasors that, even if exhaustively knowable and statable, if denied retail would make the inference unsurveyable. That is, they are not devices that *make* non-monotonic inferences monotonic. The proper term for a Latin phrase whose utterance could do *that* is '*magic spell*'. If it is thought of as a wholesale proviso covering all possible defeasors, the effect of adding '*ceteris paribus*' to the statement of the inference that if I strike this dry, well-made match, then it will light, would be to say: "unless for some reason it doesn't," or "except in those circumstances when it doesn't." That is not producing an inference that is *monotonic*; it is producing one that is *trivial*. The real expressive function of *ceteris paribus* clauses is not *magically* to *remove* the non-monotonicity of material inferences, nor to replace them with other monotonic ones, but rather *explicitly* to *acknowledge* their non-monotonicity: to mark the inference being endorsed as one that has unspecified, but potentially important defeasors.[16]

---

[16] For empirical claims involving theoretical vocabulary, this is obvious. For theoretical vocabulary is, by definition, vocabulary that can *only* correctly be applied as the conclusion of an inference. But the justification even of beliefs acquired *non*-inferentially, through observation typically will involve appealing to the reliability of the observer's differential responsive dispositions to endorse such claims

The fourth premise is that at any given time, many, if not most, of a subject's beliefs could only be justified by exhibiting them as the conclusions of material inferences. We might call a believer "epistemically responsible" insofar as she acknowledges a commitment to being able to justify many, if not most, of her beliefs, under suitable circumstances. My fifth premise is that in order to count as a discursive practitioner, one must be at least minimally epistemically responsible. Present purposes will not require that we attempt to quantify what the minimal level of such responsibility is.

Developing a theme from my previous lecture, we can draw a preliminary conclusion. The five considerations advanced so far together entail that epistemically responsible believers face a potentially intractable *updating problem*. Every change of belief, no matter how small, is *potentially* relevant to the justification of every prior belief. Acquiring a new belief means acquiring what, for any material inference the believer endorses and relies upon for justification, might possibly turn out to be a defeasor. And giving up any belief means giving up not only a premise that might previously have been relied upon in justification, but also a potential counter-defeasor (for instance, a magnetic field's not being a defeasor to the match's lighting if there is a Faraday cage inside the field).

Now, it is not practically feasible explicitly to review all of one's beliefs every time one's beliefs change, in order to check which are and which are not still justifiable. If that were what epistemic responsibility demanded, then it would be a pointless, impossible ideal. Language users who *do* not (because they *can*not) do *that*, must *practically* distinguish, among all the inferences that rationalize their current beliefs, which of them are update *candidates*, in the light of the current change of belief (let us say, for simplicity, a newly added belief). That is practically to associate with the new belief a set of material inferences of which it is a potential defeasor. The potential defeasors in this way associated with each material inference endorsed in turn define (by complementation) the range of counterfactual robustness practically associated with that inference.[17]

under a range of circumstances. The inference from my being a reliable reporter of red things in good light to my responsively elicited claim that something is red being true can be a good material inference. But it is non-monotonic, defeasible by a whole range of collateral circumstances.

[17] Somewhat more carefully put: assuming some length restriction ensuring finiteness of the set of logically non-compound sentences involved, the ability to associate with each sentence a set of

I conclude that in view of the non-monotonicity of material inference, the practical task of updating the rest of one's beliefs when some of them change is tractable in principle only if those who deploy a vocabulary practically discriminate ranges of counterfactual robustness for many of the material inferences they endorse. If that is right, then establishing the modal Kant-Sellars thesis requires further showing how to introduce modal vocabulary on the basis of such counterfactual conditionals, and how to use modal vocabulary to make those counterfactual conditionals explicit. Ryle's remarks suggest a strategy for both: treat "If *p were* true, *q would be* true," as equivalent to "It is not possible that *p* and *not-q*." In the next lecture I will show how to follow out this strategy in detail, by treating the claim that *q* follows from *p* as equivalent to the claim that everything materially incompatible with *q* is materially incompatible with *p*—so that to say that "Coda is a dog" entails "Coda is a mammal" is to say that everything incompatible with his being a mammal is incompatible with his being a dog.

4 The normative Kant-Sellars thesis

Before turning to that project of connecting material inferential relations with an implicitly modal notion of material incompatibility, however, I want to consider an analogue of the Kant-Sellars thesis about *modal* vocabulary that applies instead to *normative* vocabulary.

Kant read Hume's theoretical and practical philosophies as raising variants of a single question. On the side of theoretical reasoning, Hume asks what our warrant is for moving from descriptions of what *in fact* happens to characterizations of what *must* happen, and what could not happen. How, he wants to know, can we rationally justify the move from descriptions of matter-of-factual regularities to formulations of necessary laws? On the side of practical reasoning, Hume asks what our warrant is for moving from descriptions of how things are to prescriptions of how they ought to be. How, he wants to know, can we rationally justify the move from 'is' to 'ought'? In Kant's terminology, these are both species of 'necessity': *practical*

inferences of which it is a potential defeasor can be algorithmically elaborated into (and hence is PP-sufficient for) the ability to associate with each inference a set of potential defeasors, and hence again, the set of non-defeasors.

(including moral), and *natural* necessity, respectively. For him, 'necessary' (*notwendig*) just means "according to a *rule*." Hume's predicament is that he finds that even his best understanding of *facts* does not yield an understanding of *rules* governing and relating those facts, underwriting assessments of which of the things that actually happen (something we can experience) *must* happen (are *naturally* necessary), or *ought* to happen (are *normatively* or *practically* necessary).

As we have seen, on the modal side, Kant's response is that Hume's predicament is not a real one. One cannot in fact fully understand the descriptive, empirical employment of ordinary determinate concepts such as <u>cat</u> without at least implicitly understanding also what is made explicit by the modal concepts that articulate laws. Kant mounts a corresponding line of thought on the side of *normative* or *practical* necessity. Normative concepts make explicit commitments that are implicit in any use of concepts, whether theoretically in judgment or practically in acting intentionally—that is, in endorsing practical maxims. *Judgment* and *agency* are implicitly normative phenomena because they consist in the application of concepts, and applying concepts is undertaking commitments and responsibilities whose *content* is articulated by those concepts. (For Kant, specifically *moral* normative vocabulary makes explicit commitments that are already implicit in the practical use of concepts to endorse maxims, ends, and plans.)

I am not going to go into how Sellars builds on this thought, because I will develop it in a somewhat different way. Suffice it to say that in the light of Kant's parallel responses to Hume's parallel concerns with the credentials of modal and normative vocabulary—concerns couched in epistemological terms, but at base semantic in character—we can formulate a *normative* Kant-Sellars thesis by analogy to the *modal* one. It is the claim that in order to apply or deploy ordinary, empirical, descriptive vocabulary, including observational vocabulary—and hence, in order to deploy any autonomous vocabulary whatsoever—one must already be able to do everything needed to introduce normative vocabulary. Articulated in terms of meaning-use analysis, it is the claim that there are practices PV-necessary for engaging in any autonomous discursive practice that are PP-sufficient for practices PV-sufficient to deploy normative vocabulary. If, again by analogy to the modal case, we add the claim that normative vocabulary is VP-sufficient to *specify* those aspects of the practices that are PV-necessary for any ADP, we have the full-blown claim that normative vocabulary is elaborated-explicitating,

or LX, for all autonomous vocabularies. The by-now familiar MUD for the resultant complex meaning-use relation among vocabularies is set out in Figure 4.4.

How might one argue for the normative Kant–Sellars thesis? I have been working all along with the idea that any autonomous set of practices can be intelligible as deploying a vocabulary—that is, as being *discursive* or *linguistic* practices—only insofar as those practices attribute to some performances the pragmatic significance of *assertions*, and that it is a necessary feature of that pragmatic significance that assertions can serve both as premises and conclusions of *inferences*. The notions of <u>asserting</u> and of <u>inferring</u> are, on this account, essentially and indissolubly linked. This is to say that every autonomous *discursive* practice must include core practices of *giving and asking for reasons*. It is playing a suitable role in such a constellation of practices that makes the sign-designs whose production can have in that context the pragmatic significance of being an assertion—something that can both serve as and stand in need of a reason—qualify as *declarative sentences*. And standing in those inferential (justificatory, evidential) relations is a necessary condition of those sentences being intelligible as expressing *propositional contents*.[18]
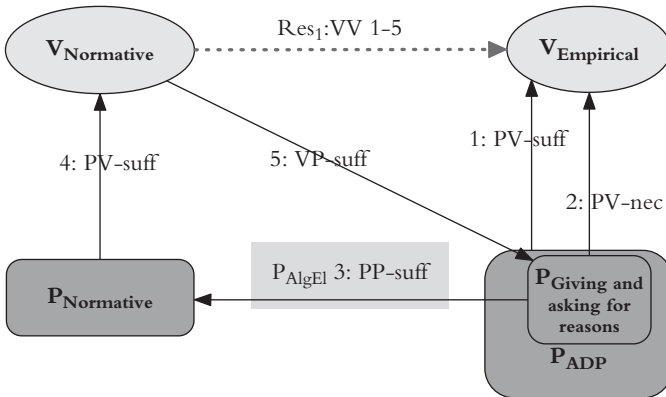
Figure 4.4 Normative Kant–Sellars thesis: normative vocabulary is elaborated-explicating (LX)

---

[18] For my purposes here I do *not* need to claim that inferential articulation, broadly construed, is *sufficient* to constitute propositional content. I need only the weaker claim that it is a *necessary* feature: that nothing that could *not* play the role of premise and conclusion of an inference could be intelligible as propositionally contentful.

It is these core practices of giving and asking for reasons that I propose as being both PV-necessary for every autonomous discursive practice (as I have just been claiming) and PP-sufficient for, in the sense of algorithmically elaboratable into, practices PV-sufficient for the introduction of normative vocabulary, which can then serve explicitly to specify key features of those practices. In particular, I will argue that no set of practices is recognizable as a game of giving and asking for *reasons* for *assertions* unless it involves implicitly (practically) acknowledging at least two sorts of normative status, *commitments* and *entitlements*, and some general structures relating them.

Suppose we have a set of counters or markers such that producing or playing one has the social significance of making an assertional move in the game. We can call such counters 'sentences'. Then, for any player at any time, there must be a way of partitioning sentences into two classes, by distinguishing somehow those that he is disposed or otherwise prepared to assert (perhaps when suitably prompted). These counters, which are distinguished by bearing the player's mark, being on his list, or being kept in his box, constitute his score. By playing a new counter, making an assertion, one alters one's own score, and perhaps that of others.

Here is my first claim: for such a game or set of toy practices to be recognizable as involving *assertions*, it must be the case that playing one counter, or otherwise adding it to one's score, can *commit* one to playing others, or adding them to one's score. If one asserts "The swatch is red," one *ought* to add to one's score also "The swatch is colored." Making the one move *obliges* one to be prepared to make the other as well. This is not to say that all players actually *do* have the dispositions they *ought* to have. One might not act as one is committed or obliged to act; one can break or fail to follow this sort of rule of the game, at least in particular cases, without thereby being expelled from the company of players of the asserting game. Still, I claim, assertional games must have rules of this sort: rules of *consequential commitment*.

Why? Because to be recognizable as *assertional*, a move must not be *idle*, it must make a *difference*, it must have *consequences* for what else it is appropriate to do, according to the rules of the game. Assertions express judgments or beliefs. Putting a sentence on one's list of judgments, putting it in one's belief box, must have consequences for how *else* one ought, rationally, to act, judge, and believe. We may be able to construct cases where it is intelligible to attribute beliefs that are consequentially inert and

isolated from their fellows: "I just believe that cows look goofy, that's all. Nothing follows from that, and I am not obliged to act in any particular way on that belief." But *all* of our beliefs could not intelligibly be understood to be like this. If putting sentences onto my list or into my box *never* has consequences for what else belongs there, then we ought not to understand the list as consisting of my *judgments*, or the box as containing my *beliefs*.

Understanding a claim, the significance of an assertional move, requires understanding at least *some* of its consequences, knowing what *else* (what other moves) one would be committing oneself to by making that claim. A parrot, we can imagine, can produce an utterance perceptually indistinguishable from an assertion of "That's red." Our nonetheless not taking it to have asserted that sentence, not to have made a move in that game, *is* our taking it that, unaware as it is of the inferential involvements of the claim that it would be expressing, of what it would be committing itself to were it to make the claim, it has not thereby succeeded in committing itself to anything. Making that assertion is committing oneself to such consequences as that the swatch is colored, that it is not green, and so on.

For this reason we can understand making a claim as taking up a particular sort of normative stance towards an inferentially articulated content. It is *endorsing* it, taking *responsibility* for it, *committing* oneself to it. The difference between treating something as a claiming and treating it just as a brute sounding off, between treating it as making a move in the assertional game and treating it as an idle performance, is just whether one treats it as the undertaking of a commitment that is suitably articulated by its consequential relations to other commitments. These are *rational* relations, whereby undertaking one commitment *rationally* obliges one to undertake others, related to it as its inferential consequences. These relations at least partly articulate the *content* of the commitment or responsibility one undertakes by asserting a sentence. Apart from such relations, there is no such content, hence no assertion.

The next claim I want to make is that practices incorporating a game of giving and asking for reasons must involve acknowledgment of a *second* kind of normative status. I have said that making a move in the assertional game should be understood as acknowledging a certain sort of *commitment*, articulated by consequential inferential relations linking the asserted sentence to other sentences. But players of the game of giving and asking for reasons must also distinguish, among the commitments an

interlocutor undertakes, a distinguished sub-class to which she is *entitled*. *Linguistic rationalism* understands *assertions*, the fundamental sort of speech act, as essentially things that can both serve as and stand in need of *reasons*. Giving reasons for a claim is producing other assertions that *license* or *entitle* one to it, that *justify* it. Asking for reasons for a claim is asking for its *warrant*, for what *entitles* one to that commitment. Such a practice presupposes a distinction between assertional commitments to which one *is* entitled and those to which one is *not* entitled. Reason-giving practices make sense only if there can be an issue as to whether or not practitioners are entitled to their commitments.

Indeed, I take it that liability to demands for *justification*, that is, demonstration of *entitlement*, is a major dimension of the *responsibility* one undertakes, the *commitment* one makes, in asserting something. In making an assertion one implicitly acknowledges the propriety, at least under some circumstances, of demands for reasons, for justification of the claim one has endorsed, the commitment one has undertaken. Besides the *committive* dimension of assertional practice, there is the *critical* dimension: the aspect of the practice in which the propriety of those commitments is assessed. Apart from this critical dimension, the notion of *reasons* gets no grip.

So the overall claim is that the sense of endorsement that determines the force of assertional speech acts involves, at a minimum, a kind of *commitment* the speaker's *entitlement* to which is always potentially at issue. The assertible contents expressed by declarative sentences whose utterance can have this sort of force must accordingly be inferentially articulated along both normative dimensions. *Downstream*, they must have inferential *consequences*, commitment to which is entailed by commitment to the original content. *Upstream*, they must have inferential *antecedents*, relations to contents that can serve as premises from which entitlement to the original content can be inherited.

5 Conclusion

If that is right, then discursive practitioners as such must be able in practice to take or treat each other and themselves as exhibiting *normative* statuses: as being *committed* and *entitled* to contents expressed by the declarative

sentences whose free-standing utterance has the pragmatic significance of acknowledging commitments and claiming entitlements. Since, by hypothesis, these practitioners can already make assertions, the introduction of normative vocabulary permitting one explicitly to *say that* someone is committed or entitled to a claim requires only that new vocabulary, "S is committed to *p*," and "S is entitled to *p*," be deployed with the *circumstances* of application that one can assert these sentences formed using the new normative vocabulary whenever one would in practice respond to S as having the commitment or entitlement labeled with the sentence *p*, and with the *consequences* of application that whenever one asserts one of these new normative sentences, one must also take or treat S in practice as having the corresponding commitment or entitlement. Introducing vocabulary playing this role requires only the algorithmic elaborative abilities I have called "*response substitution*" (along with the arbitrary formation and permutation of states), together with the sort of basic deontic scorekeeping abilities I have argued one must possess in order to engage in practices of giving and asking for reasons at all. (Compare the instructions for introducing *conditionals* that I offered in Lecture 2.) Further, when used with these circumstances and consequences of application, it is clear that when one of these new normative sentences is asserted, the pragmatic significance of that speech act will be to *say that* someone is committed or entitled to a claim, making propositionally *explicit* a practical attitude—taking or treating someone in practice *as* committed or entitled to a claim—that before the advent of the new vocabulary remained *implicit* in what one *did*.

My overall claim is that both the modal and the normative Kant-Sellars theses are true. In order to be able to talk at all, to make claims and inferences, one must already know how to do everything necessary in principle (in the precise sense of 'in principle' given by the notion of algorithmic elaboration) to deploy alethic modal and deontic normative vocabulary. If so, one cannot be stuck in the position Hume took himself to be in: understanding ordinary empirical, descriptive vocabulary, but with that providing no grip on the use of modal and normative vocabulary. The *semantic* relations between what is expressed by the use of empirical descriptive vocabulary, on the one hand, and what is expressed by the use of modal and what (something different) is expressed by normative vocabulary, on the other, are essentially *pragmatically mediated* ones. To

understand the relation between how things merely *are* and how they *must* be or (a different matter) *ought* to be, one must look at what one is *doing* in *saying* how things are, and what is required to *say* what one is thereby doing. Transposing Kant's response to Hume into this pragmatist key requires the metaconceptual resources of meaning-use analysis, which is what enables us to be clear about the pragmatically mediated semantic relations on which that response depends.

Coming to understand both modal and normative vocabulary as standing in the complex resultant pragmatically mediated semantic relation of being LX to—elaborated from and explicating of—practices integral to every autonomous discursive practice will turn out also to be the key to understanding a deep and illuminating feature of the relation of these two vocabularies, not just to vocabulary use in general, but also to each other. It supplies the raw materials for filling out and developing Sellars's suggestive claim that modal vocabulary is a 'transposed' language of norms. In the next lecture I will begin to explore the relations between normative and modal vocabulary that become visible in this way, showing how normative vocabulary can serve both as a pragmatic metavocabulary for modal vocabulary and as the basis for a directly modal formal semantics for ordinary empirical vocabulary that does not appeal in any way to a notion of <u>truth</u>. In the final lecture, that discussion will be brought together with the discussion of modality and normativity from the two lectures that precede it, culminating in an understanding of *discursive intentionality*, the ultimately semantic relations between knowing subjects and their cognitive objects that is expressed by *intentional* vocabulary, in terms of the relations between what is expressed by *normative* and *modal* vocabularies.

5

Incompatibility, Modal Semantics, and Intrinsic Logic

1 Introduction

I closed Lecture 4 with an argument building on the idea that every autonomous discursive practice, in order to count as a *discursive* or *linguistic* practice, in order to count as deploying any *vocabulary*, must include performances that have the *pragmatic* significance of *assertions*, which on the *syntactic* side are utterances of *declarative sentences*, and whose *semantic* content consists of *propositions*. These pragmatic, syntactic, and semantic conditions form an indissoluble package, in the sense that one cannot properly understand any of the concepts <u>assertion</u>, <u>sentence</u>, and <u>proposition</u> apart from their relation to each other. This is the *iron triangle of discursiveness* (Figure 5.1).

I then proceeded to look at the pragmatic presuppositions of the assertional practices that are, on this account, PV-necessary to deploy any autonomous vocabulary. Here my claim was that no set of practices could count as according some performances the pragmatic significance of assertions unless it includes practices of giving and asking for *reasons*. That
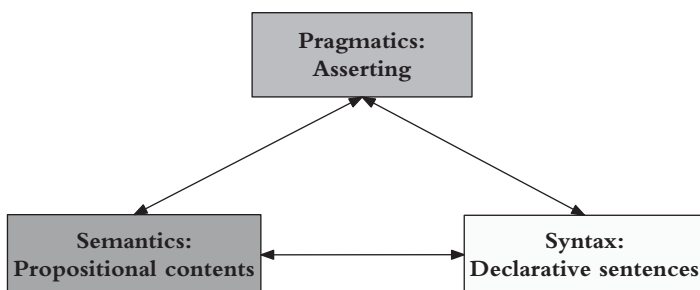
Figure 5.1 The iron triangle of discursiveness

is the claim that within the pragmatic dimension of the triad, *asserting* and *inferring* also form an indissoluble package, each element of which is in principle intelligible only in a context that includes the other. Assertional and inferential practices are reciprocally PP-necessary.

I then argued that any constellation of social practices is intelligible in principle as including the giving and asking for *reasons*—making *claims* whose status depends on their *inferential* relations to other claims that are their consequences, or have them as their consequences, or rule them out—only if it includes the capacity to distinguish two sorts of normative status as part of the pragmatic significance practically attributed to a speech act. To be giving and asking for *reasons*, interlocutors must practically distinguish (be able to respond differentially to) the sentences to which their interlocutors and they themselves are *committed* (based on those they are disposed to assert). And they must distinguish the sentences to which their interlocutors and they themselves are *entitled* (based on those they are committed to). These practical discriminative capacities need not be infallible (by any standard of ultimate correctness), and they need not be complete. But unless interlocutors make at least these two sorts of discrimination, what they are doing does not deserve to count as producing and consuming *reasons*, hence not as practically according some performances the pragmatic significance of *assertions*, hence not as deploying any autonomous *vocabulary*.[1]

My interest in the previous lecture was in arguing that these practices-or-abilities to discriminate commitments and entitlements are, in the terms of the sort of meaning–use analysis I have been developing here:

- PV-necessary for deploying any autonomous vocabulary,
- PP-sufficient by algorithmic elaboration for engaging in practices that are

---

[1] These are strong claims, no doubt contentious because tendentious: framed from the point of view of a normative pragmatist rationalism about the discursive. Those who are not convinced, those not tempted, and those not even willing to suspend disbelief on these points should just consider the remarks that follow as restricted to that sub-class of discursive practices that *does* exhibit the structure being considered—a sub-class that should at least be admitted to be large and significant, even by those who doubt that it plays the foundational and demarcational role here attributed to it. In any case, the principal arguments and constructions to be presented here as articulating incompatibility semantics do not depend on the particulars of the normative pragmatic metavocabulary in terms of which I want to understand <u>incompatibility</u>.

- PV-sufficient to deploy normative vocabulary, which is
- VP-sufficient to specify those original universally PV-necessary prac-tices-or-abilities.

In sum, it was to argue that normative vocabulary—paradigmatically 'commitment' and 'entitlement'—stands in the complex resultant meaning-use relation of being elaborated-explicating (LX) with respect to every autonomous vocabulary. Whatever the status of that argument may be, my purpose here is to consider a different complex resultant meaning-use relation that the explicitly normative vocabulary of commitment and entitlement stands in to other vocabularies of philosophical interest, prin-cipal among them being alethic *modal* vocabulary. The relation I will focus on is that of one vocabulary's being a *pragmatic metavocabulary* for another. I want to explore a particular construction according to which normative vocabulary can serve as a pragmatic metavocabulary for *logical* vocabulary, including *modal* vocabulary, and how in those terms it can be seen to serve as such a metavocabulary for *semantic* vocabulary more generally. Along the way we will learn some lessons about logic and modality, and especially about the relation of *truth* and *compositionality* to semantics, that I think are of general interest, quite apart from the way in which they emerge from the particular pragmatic-analytic project I am pursuing here.

2 Incompatibility

The story I told about how engaging in practices of giving and asking for reasons requires the practical differential responsive ability to take or treat someone as committed and as entitled to the claims expressed by various sentences lets us make sense straightaway of two sorts of inferential relations between propositional contents on the semantic side, and corresponding practical dispositions on the pragmatic side. One takes or treats q as an inferential consequence of p in one sense by being disposed to attribute *commitment* to (what is expressed by) q to whomever one credits with commitment to (what is expressed by) p. And one takes or treats q as an inferential consequence of p in another sense by being disposed to attribute *entitlement* to the claim that q to whomever one credits with

entitlement to the claim that *p*.² The first sort, commitment-preserving inferential relations, is a generalization, to include the case of non-logical, material inferences, of obligatory, *deductive* inferential relations. The second sort, entitlement-preserving inferential relations, is a generalization, to include the case of non-logical, material inferences, of permissive, *inductive* inferential relations. For example, anyone who is committed to a plane figure being rectangular is committed to its being polygonal. And the old nautical meteorological homily, ''Red sky at night, sailor's delight; red sky in morning, sailor take warning,'' tells us that anyone who sees a colorful sunrise is entitled to the claim that a storm that day is probable. But here the reasoning is only probative, not dispositive. The colorful sunrise provides *some* reason to predict a storm, but does not yet settle the matter. Other considerations, such as a rising barometer, may license one not to draw the conclusion one would otherwise be entitled to by the original evidence.

The abilities to take or treat interlocutors (including oneself) as committed or entitled to propositional contents expressed by various declarative sentences are PP-sufficient for the practical responsive recognition of another sort of semantic relation among propositional contents. This is because being disposed to respond to anyone who is *committed* to *p* as thereby precluded from counting as *entitled* to *q* (and vice versa) is treating *p* and *q* as *incompatible*. On the pragmatic side, this is a *normative* relation. It is not that one *can*not undertake incompatible commitments, make incompatible assertions. Finding that one has done so is an all-too-common occurrence. But the effect of doing so is to alter one's normative status: to undercut any entitlement one might otherwise have had to either of the incompatible commitments, for each commitment counts as a decisive reason against entitlement to the other, incompatible one.

On the pragmatic side, incompatibility can accordingly be thought of as a consequential relation like the other two:

- Underline Incompatibility of *p* and *q*: If *S* is committed to *p*, then *S* is *not* entitled to *q*.
- Underline Committive consequence: If *S* is committed to *p*, then *S* is committed to *q*.

---

² As will appear, entitlement-preserving inferences are always defeasible; the entitlement one acquires thereby is only *prima facie*. One is not entitled to the conclusion of a good entitlement-preserving inference if one is committed to something incompatible with it.

- <u>Permissive consequence</u>: If *S* is committed and entitled to *p*, then *S* is (*prima facie*) entitled to *q*.

But it is not immediately an *inferential* relation, since the conclusion is the *withholding* of a primary normative status, rather than the *inheritance* of one. Incompatibility relations *do*, however, underwrite a kind of inferential relation. The idea is an old one. Sextus Empiricus says, perhaps referring to Chrysippus:

And those who introduce the notion of <u>connexion</u> say that a conditional is sound when the contradictory of its consequent is incompatible with its antecedent.[3]

My concern is not with when a *conditional* is sound, but with when the underlying *inference* that such a conditional is VP-sufficient to specify is a good one, in the material (that is, non-, or better, pre-logical) sense of "good inference" I am trying to articulate. And I do not want to assume at this stage that we are in a position to identify the *contradictory* of any claim. But the notion of <u>material incompatibility</u> can serve in its place. Making those adjustments yields the following definition:

> *p incompatibility-entails q* just in case everything incompatible with *q* is incompatible with *p*.

Thus "Pedro is a donkey," incompatibility-entails "Pedro is a mammal," for everything incompatible with Pedro's being a mammal (for instance, Pedro's being an invertebrate, an electronic apparatus, a prime number) is incompatible with Pedro's being a donkey.

I said before that the inferential relations among the propositional contents expressed by declarative sentences that correspond on the semantic side to inheritance of *commitment* can be thought of as a generalization (to the material case) of *deductive* inferential relations, and that those corresponding to inheritance of *entitlement* can be thought of as a generalization to the material case of *inductive* inferential relations. So we may ask: do *incompatibility*-entailments similarly generalize some kind of inferential relation that we already recognize in other terms? I think that they do, and that the inferences in question are counterfactual-supporting, *modally* robust inferential relations: the kind of inferences made explicit by *modally*

---

[3] Sextus, *Pyrrhoneiae Hypotyposes*, ii, 110−12, in William Kneale and Martha Kneale, *The Development of Logic* (Oxford University Press, 1962), 129.

qualified conditionals. The fact that the *properties* of being a donkey and being a mammal stand in the relation of incompatibility-entailment means that every *property* incompatible with being a mammal is incompatible with being a donkey. If two properties (such as being a mammal and being an invertebrate) are incompatible then it is *impossible* for any object simultaneously to exhibit both. And that means that it is *impossible* for anything to be a donkey and not be a mammal. That is why the incompatibility-entailment in question supports counterfactuals such as "If my first pet (in fact, let us suppose, a fish) *had been* a donkey, it *would have been* a mammal." We could say: "Necessarily, anything that is a donkey is a mammal."

On the semantic side, <u>incompatibility</u> is an implicitly *modal* notion. On the pragmatic side, the *normative* concepts of <u>commitment</u> and <u>entitlement</u> provide a pragmatic metavocabulary VP-sufficient to *specify* practices PV-sufficient to *deploy* that modal notion. That is, they let us *say* what it is one must *do* in order thereby to be taking or treating two claims *as* incompatible.[4] To begin to explore the consequences of this pragmatically mediated semantic relation between normative and modal vocabularies, we may consider the sort of grip on the semantics of expressions—the meanings expressed by deploying vocabularies—that one gets by thinking of their contents in terms of incompatibilities. I argued in Lecture 4 that there is an intimate connection between the conceptual contents expressed by vocabularies and the counterfactually robust inferences they are involved in. We might hope that a semantic metavocabulary centered on *incompatibility* would have the right expressive resources to make explicit important features of such contents. One case where we have particularly clear criteria of adequacy for our semantics is *logical* vocabulary. So I will be specifically concerned to offer an incompatibility semantics for logical vocabulary. Again, since <u>incompatibility</u> is at least implicitly itself a *modal* notion, we will want to see what an incompatibility semantics for modal vocabulary might look like. On this basis, one would hope to continue by elaborating a modal intensional semantics for *non*-logical vocabulary, as was done with possible worlds semantics in the second phase of the modal revolution.

---

[4] In my final lecture I shall be concerned to explore in much further detail this relation between what is expressed by modal and by normative vocabulary, as a way of thinking about the intentional nexus beween objects and the subjects who make claims about and act upon them.

3 Incompatibility semantics

Here is a semantic suggestion: represent the propositional content expressed by a sentence with the set of sentences that express propositions incompatible with it.[5] More generally, we can associate with each *set* of sentences, as its semantic interpretant, the set of *sets* of sentences that are incompatible with it.[6] The generalization from seeing incompatibility as a relation among sentences to seeing it as a relation among *sets* of sentences acknowledges an important structural fact about incompatibility: one claim can be incompatible with a set of other claims without being incompatible with any of its members. On the formal, logical side, where *incompatibility* is just *inconsistency*, p is incompatible with the set consisting of $p \to q$ and $\sim q$, but not with either individually. And on the side of non-logical content, the claim that the piece of fruit in my hand is a blackberry is incompatible with the *two* claims that it is red and that it is ripe, though not with either individually—in keeping with the childhood slogan that blackberries are red when they're green.

Aiming at maximal generality, I will impose only two conditions on the incompatibility relations whose suitability as semantic primitives I will be exploring here. First, I will only consider *symmetric* incompatibility relations. This is an intuitive condition because it is satisfied by familiar families of incompatible properties: colors, shapes, quantities, biological classifications, and so on. Second, if one set of claims is incompatible with another, so too is any larger set containing it. That is, one cannot remove or repair an incompatibility by throwing in some further claims. I call this the 'persistence' of incompatibility. If the fact that the monochromatic patch is blue is incompatible with its being red, then it is incompatible with its being red and triangular, or its being red *and* grass being green.

---

[5] Since, as has just been emphasized, incompatibility relations are only *one* dimension of inferential articulation, this semantic representation of conceptual content will necessarily be only partial.

[6] This generalization opens up a number of possibilities for correspondingly generalizing the incompatibility entailment relation. One very natural way to do that is to take it that a set of sentences X incompatibility entails a set Y just in case every set Z that is incompatible with Y is incompatible with X. In this case, $X \models \{y_1, \ldots, y_n\}$ has the meaning, X entails (y_1 and … and y_n). It turns out to be more formally convenient if instead one requires that X incompatibility entails Y in case every set Z incompatible with *every sentence in Y* is incompatible with X. In this case, $X \models \{y_1, \ldots, y_n\}$ has the meaning, X entails (y_1 or … or y_n).

Given any set of sentences, we can then define a standard *incompatibility interpretation* over that vocabulary as an incoherence partition of its power set that satisfies persistence. (Two sets of sentences are incompatible if and only if their union is incoherent.) Each such incompatibility interpretation induces an *incompatibility consequence* (or entailment) *relation* $\models$ in the way already indicated: being a cat entails being a mammal in this sense because every set of properties incompatible with being a mammal is also incompatible with being a cat.

The proposal here is to use incompatibility (itself introduced by a normative pragmatic metavocabulary) as the basic element of the *semantic metavocabulary*—and not just for *logical* expressions, but for ordinary *non*-logical vocabulary as well. The semantic interpretant of an object-vocabulary sentence is taken to be the set of sets of sentences materially incompatible with it.

The result is a *modal* semantics. For *incompatibility* is a *modal* notion. Now the development of modal semantic metavocabularies—in particular, the extension of possible world semantics from its initial home as a semantics for modal *logical* vocabulary to a modal semantics for ordinary, non-logical expressions in general—is perhaps the principal technical philosophical advance of the past forty years.[7] (It is the second of three sequential modal revolutions in recent philosophy—or of three phases of one complex, multi-stage revolution—the first being Kripke's formal possible worlds semantics for modal *logic*, and the third beginning with his application of that apparatus to the semantics of proper names.) I want to take that hint, but to apply modal vocabulary to semantic projects in a somewhat different way: using the notion of <u>incompatibility</u> to provide a *directly* modal semantics. By that I mean one that does not approach modality by beginning with a more basic semantic notion of *truth*.

Classical possible-worlds semantics proceeds in two stages. Like more traditional semantics, its basic semantic notion is that of truth. It begins by relativizing evaluations of truth to points of evaluation—paradigmatically, possible worlds. Then, at the second stage, necessity and possibility can be introduced by quantification over such points of truth-evaluation—possibly exploiting structural relations among them, such as accessibility relations

---

[7] For an example, one can consider the story about the semantic distinction between attributive and non-attributive adverbs that I tell in the Afterword.

among possible worlds, or the ordering of time and place co-ordinates. The semantic interpretants of expressions are in the first instance functions from points of evaluation to extensions or truth-values. This is one natural way to capture the element of *generality* that Ryle insisted was present in all endorsements of inferences:

… some kind of openness, variableness, or satisfiability characterizes all hypothetical statements alike, whether they are recognized "variable hypotheticals" like "For all *x*, if *x* is a man, *x* is mortal" or are highly determinate hypotheticals like "If today is Monday, tomorrow is Tuesday."[8]

By contrast to such two-stage approaches, semantics done in terms of incompatibility is *directly* modal. One may, if one likes, think of the incompatibility of *p* and *q* as the impossibility of both being *true*. But that characterization in terms of truth is entirely optional. Incompatibility is itself already a modal notion, and for semantic purposes we can treat it as primitive. The explication I have offered is in pragmatic terms: *saying* (in terms of the normative notions of commitment and entitlement) what one must *do* in order to be taking or treating two claims *as* incompatible. The element of generality comes in because in assessing entailments we look at *all* the claims that are incompatible with the conclusions and the premises. One claim is an incompatibility consequence of another only if there is *no* set of sentences incompatible with the conclusion and not with the premises. And here it is important that the potential defeasors are *not* limited to sentences that are *true*. Even if as a matter of fact all the coins in my pocket are copper, that a coin is in my pocket does not *entail* that it is copper, since "This coin is silver" is incompatible with its being copper, but *not* with its being in my pocket, even though it is not *true* that it is in my pocket. For, as we want to say, it *could* be in my pocket: that non-actual state of affairs is *possible*. That modal fact is reflected in the fact that a coin's being silver is not *incompatible* with its being in my pocket. The idea that I want to explore is that once we have properly learned the lesson that modality matters in semantics because counterfactually robust inferences are an essential aspect of the articulation of the conceptual contents of sentences, the way is opened up to a *directly* modal semantics, which does *not* make what now appears as an unnecessary preliminary *detour* through assessments of *truth*.

---

[8] Gilbert Ryle " 'If', 'So', and 'Because'," in Max Black (ed.), *Philosophical Analysis* (Prentice Hall, 1950), 302–18, at 311.

This is all very abstract. In order to see incompatibility semantics in action, we should look to the case where the criteria of adequacy of a semantics are clearest: namely, to semantics for *logical* vocabulary. That, after all, is where possible-worlds semantics cut its teeth.

4   Introducing logical operators

The notion of <u>incompatibility</u> can be thought of as a sort of conceptual vector-product of a *negative* component and a *modal* component. It is <u>*non-com*possibility</u>. To use this semantic notion to introduce a negation operator into the object vocabulary, we must somehow isolate and express explicitly that negative component. The general semantic model we are working with represents the content expressed by a sentence by the set of sets of sentences incompatible with it. So what we are looking for is a way of computing what is incompatible with negated sentences (and, more generally, with sets of sentences containing them). Since we do not have any sort of yes/no evaluation of sentences in the picture (not even a relativized one), we cannot approach negation as a kind of reversal of semantic polarity. How else might we think about it?

Incompatible sentences are Aristotelian *contraries*. A sentence and its negation are *contradictories*. What is the relation between these? Well, the contradictory is *a* contrary: any sentence is incompatible with its negation. What distinguishes the contradictory of a sentence from all the rest of its contraries? The contradictory is the *minimal* contrary: the one that is entailed by all the rest. Thus every contrary of "Plane figure f is a circle"—for instance "f is a triangle," "f is an octagon," and so on—entails "f is *not* a circle." *Blue, green, yellow* all entail *not-red*. For any sentence p we are assuming that we can already pick out its contraries, that is, the (sets of) sentences that are incompatible with it. And we already have an entailment relation, defined wholly in terms of incompatibility. So we have all the resources needed to say that some other sentence q is the negation of p just in case q is the *minimal* incompatible of p: the one entailed by everything else incompatible with it.

It might happen that in some standard interpretation of the vocabulary to which p belongs, there already is such a q. But we cannot count on every sentence already having such a negation in every interpretation. So

we need to introduce new sentences, of the form N*p*, on the basis of this relation. Inspection of the definition of incompatibility entailment yields the result that N*p* will be an inferentially minimal incompatible of *p* if and only if a set of sentences is incompatible with it just in case that set entails *p*. This is equivalent to saying that what is incompatible with the negation of *p* is what is incompatible with *every* set of sentences incompatible with *p*—that is, that the incompatibility set of N*p* is just the *intersection* of the incompatibility sets of everything incompatible with *p*.[9]

This definition lets us recursively add, for every sentence *p* of the language, its negation N*p*, and to compute the incompatibility sets of those negations so as to satisfy the principle that everything incompatible with *p* entails N*p*. Extending the incompatibility relation to apply to sets of sentences that include arbitrarily iterated negations automatically extends the incompatibility consequence relation, which is defined in terms of it. And it is easy to show that that extension is *inferentially conservative*—that is, that the extended consequence relation does not add or subtract any consequences that involve only the old vocabulary.

What are the properties of negation, given this incompatibility semantics? It turns out to have all the familiar and desirable properties we expect in a negation:

- Because *p* and N*p* are guaranteed to be incompatible, every set of sentences that contains or entails both—what we are now in a position to characterize as the *inconsistent* sets of sentences—is guaranteed to be incoherent.
- Negation contraposes appropriately with incompatibility entailment. That is, *p* entails *q* if and only if *not-q* entails *not-p*.
- And every sentence is incompatibility-equivalent to its double negation: $p \models NNp$ and $NNp \models p$.

Further logical properties of negation depend on its interaction with other connectives, and accordingly must be considered after we have introduced them.

So the procedure is to start with a *material* incompatibility-and-consequence structure that articulates the contents of *non*-logical vocabulary, and on that basis introduce *logical* vocabulary—in this

---

[9] One may consult the appendices to this lecture to see how this intuition is extended to the case of sets of sentences containing negations.

case negation—whose content is derived from that of the non-logical vocabulary on which it is based. A corresponding procedure permits the introduction of *conjunction*. Here the most important fact to acknowledge is that something can be incompatible with a conjunction even though it is not incompatible with either conjunct. That the fruit in my hand is a blackberry is incompatible with its being red *and* ripe, even though it is *not* incompatible with either one individually. This is the phenomenon that led us to think about incompatibility relations among *sets* in the first place. And that is the clue as to how to compute the incompatibilities of conjunctions. What is incompatible with the conjunction Kpq should just be whatever is incompatible with the *set* $\{p, q\}$. Once again we can introduce conjunctions recursively and conservatively in this way, along with negations, so as to extend any standard incompatibility relation by computing incompatibilities for all sentences formed from basic vocabulary of primitive proposition letters by arbitrary iterations of conjunction and negation.

It is easy to show that under this definition, conjunction acts like conjunction. It is obvious from the semantic definition that the set consisting of the two premises p and q entails their conjunction ($p, q \models K pq$), and it follows immediately from the persistence of incompatibility that a conjunction entails each of the conjuncts (K$pq \models p$ and K$pq \models q$). It is less obvious, but turns out to be true, that this definition also validates the principle that if p entails q and p entails r, then p entails their conjunction.

In fact, conjunction behaves classically. Furthermore, it interacts with negation in the familiar ways: the full logic is distributive. To make a long story short, the logic generated by these semantic definitions of negation and conjunction in terms of incompatibility is just classical logic. Notice that negation and conjunction are not interpreted semantically by anything at all like *truth*-functions. As we'll see, their semantics is in both cases *intensional* in *definition*, but nonetheless *extensional* in *result*—in the sense that it yields just the theorems of classical two-valued logic.

What sets incompatibility semantics apart, however, is that we can exploit the fact that *incompatibility* is a *modal* semantic primitive to introduce *modal* logical vocabulary in the very same setting, and the very same terms, in which we introduce the classical *non*-modal logical vocabulary.

On the semantic approach I am pursuing, to introduce a connective one specifies how to compute its incompatibilities. So the question is: what intuitively should be taken to be *incompatible* with *necessarily-p*, that is, with the *necessity* of *p*? Put otherwise, what claims rule out the *necessity* of *p*? Clearly, anything incompatible with *p* is incompatible with *necessarily-p*. Given the definition of entailment, this just says that the rules for computing the incompatibilities of *necessarily-p* should ensure that *necessarily-p* entails *p*. But what else is incompatible with the *necessity* of *p*, besides the things that are incompatible with *p*? Here is the basic thought. **To be incompatible with *necessarily-p* is to be** (self–incompatible or) **compatible with something that *does not entail p*.** For anything compatible with something that does not entail *p* is compatible with something that does not necessitate *p*, and so leaves open the possibility that *p* is not necessary.

A similar line of thought applies to *possibility* in relation to incompatibility, permitting us to introduce *possibly-p* as well as *necessarily-p*. Whatever is incompatible with *possibly-p* should be incompatible with *p* (ensuring that *p* entails *possibly-p*). But only some things that rule out *p* also rule out the *possibility* of *p*. Which are those? Here is an idea. **To be incompatible with *possibly-p* is to be incompatible with *everything* that is *compatible* with something *compatible* with p.** For anything compatible with something compatible with *p* is compatible with something that leaves the *possibility* of *p* open. It turns out to be straightforward to show that, according to these definitions, *possibly-p* is incompatibility–equivalent to *not-necessarily-not-p*, and *necessarily-p* is incompatibility–equivalent to *not-possibly-not-p*, given the way we have defined negation above. So these definitions fit together in the way we would expect.

To make another long story short, the modal–logical theorems that are valid on all standard incompatibility frames given these definitions are just those of the familiar Lewis system S5. This is the system in which it is true both that whatever is *necessary* is *necessarily* necessary and that whatever is *possible* is *necessarily* possible. In the usual Kripke semantics, this is the modal logic generated by accessibility relations among possible worlds that are reflexive, symmetric, and transitive. In the tangled jungle of modal–logical systems, this is the unexciting, well–studied, well–behaved, plain–vanilla modal analogue of the classical non–modal propositional calculus.

5 Meaning-use analysis

Figure 5.2 shows a meaning-use diagram corresponding to this incompatibility semantics for modal logical vocabulary.

Here is some help in reading it:

- Basic meaning-use relations (MURs) 1–3 are by now familiar. I have argued that every autonomous discursive practice must include practices of giving and asking for reasons—as part of the iron triangle of discursiveness—and that that involves distinguishing in practice between the deontic statuses of commitment and entitlement.
- We saw last time that that is sufficient to introduce normative vocabulary, specifically the deontic vocabulary of 'commitment' and 'entitlement', which is VP-sufficient to specify the triadic inferential
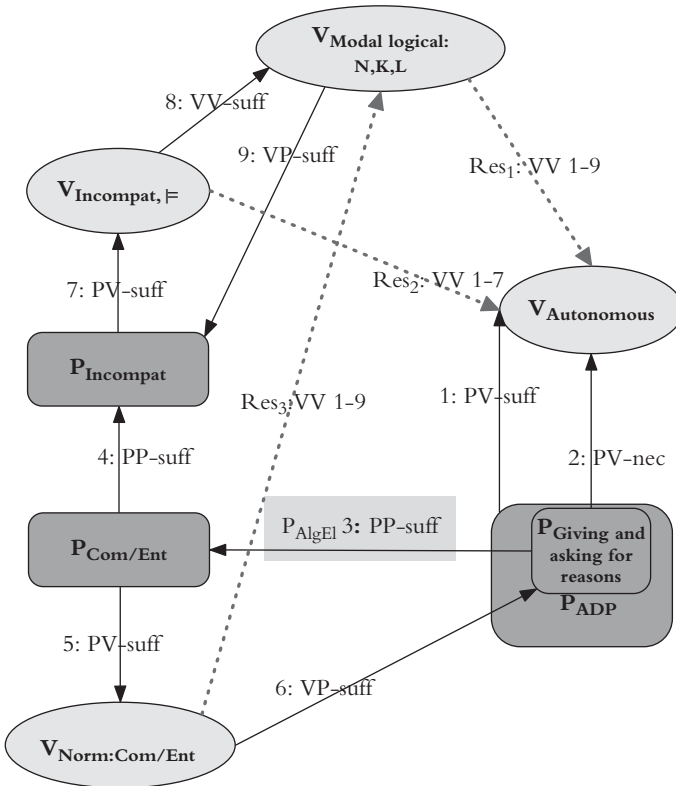
Figure 5.2 Incompatibility semantics for modal logic

substructure of practices of giving and asking for reasons. Those facts are represented by MURs 5 and 6.

- We also saw how practically distinguishing commitments and entitlements underwrites a notion of practical incompatibility of commitments, where commitment to one claim is taken or treated as sufficient to rule out entitlement to another. That is MUR 4, which permits the introduction of a semantic metavocabulary that lets one say that two claims are incompatible, and that claims stand in the relation of incompatibility-entailment, which is MUR 7.

- We have now seen how that semantic metavocabulary allows one to extend the original vocabulary by introducing modal-logical vocabulary (MUR 8), which has the expressive power to define a connective that says in that object-vocabulary *that* two claims are incompatible: LNK*pq*. Basic MUR 9 accordingly exhibits modal-logical vocabulary as a kind of *semantic* metavocabulary for incompatibility.

- Complex resultant MUR Res_1 analyzes the sense in which the vocabulary of modal logic S5 is *implicit* in the use of any autonomous vocabulary. This analysis is a further cashing-out of what last time I called "The modal Kant-Sellars thesis."

- Complex resultant MUR Res_2 codifies an analysis of the possibility of using incompatibility and incompatibility-entailment as a semantic metavocabulary for any autonomous vocabulary.

- Finally, complex resultant MUR Res_3 represents a new relation between the *normative* vocabulary of commitment and entitlement and the *modal* vocabulary of necessity and possibility. It represents a detailed analysis of a sense in which we could understand Sellars's dictum that "the language of modality is a 'transposed' language of norms." When I introduced that slogan in Lecture 4, I suggested that the way to fill in Sellars's black-box notion of 'transposition' was in terms of the pragmatically mediated semantic relation of providing a *pragmatic* metavocabulary. I offered a simple MUD as a representation of this relation (repeated here as Figure 5.3).

We are now in a position to unpack what were there represented as two basic MURs. The PV-sufficiency of a set of modal practices for the deployment of a modal vocabulary in this simple diagram corresponds to the complex MUR that is the resultant of basic MURs 7 and 8 in the MUD
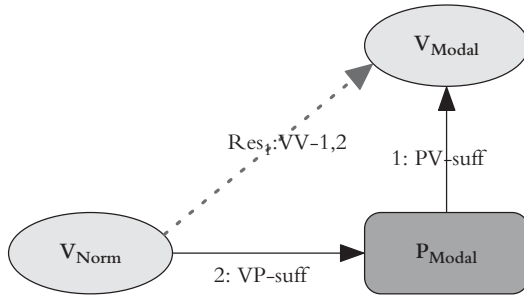
Figure 5.3 "The language of modalities is a 'transposed' language of norms."

for modal logic (Figure 5.2). The VP-sufficiency of a normative vocabulary to specify those implicitly modal vocabulary-deploying practices now shows up as the complex MUR that is the resultant of relations 3, 4, 5, and 6. In fact, we ought to include the other basic MURs that occur in our diagram, and identify the resultant representing the fact that normative vocabulary can serve as a pragmatic metavocabulary for modal vocabulary in the simple MUD with the complex resultant relation 3 in the more complex MUD.

The MUD of Figure 5.2 accordingly offers a detailed analytic reading of the Sellarsian claim that "the language of modality is a 'transposed' language of norms," understood as asserting a complex pragmatically mediated semantic relation between deontic and alethic modal vocabularies. Now, Sellars's claim might or might not be correct. And this interpretation-as-analysis of it might or might not be correct. But I take it to be a signal measure of the power of the metaconceptual apparatus of meaning-use analysis that it so much as permits the expression of this detailed a reading. And I have worked hard here to justify MURs 4, 7, 8, and 9. Regardless of how successful those efforts have been, the fact that the meaning-use analysis tells us *exactly* what constellation of sub-claims we must argue for in order to justify the overall account seems to me to constitute concrete progress in our grasp of and control over the philosophical claims we make in this area.

The complex pragmatically mediated semantic relation between deontic and alethic modal vocabularies that shows up here indicates that there is a deep relation between what in the previous lecture I called the "*modal* and *normative* Kant-Sellars theses." In the final lecture I will have more to say about this relation, and about what it has to do with what is expressed by *intentional* vocabulary.

6 Semantic holism: recursive projectibility without compositionality

Returning closer to ground-level, however, there are two more lessons I take to be of some potential philosophical significance that can be drawn from the construction of an incompatibility semantics for modal-logical vocabulary. The first concerns debates about semantic *holism* and *compositionality*.

As with the familiar Kripke semantics for modal vocabulary, the meta-vocabulary in which incompatibility semantics is conducted is entirely *extensional*. The semantic interpretants of sentences (and theories) are just sets (of sets of sentences), and the semantic interpretants of logically compound sentences are computed by purely set-theoretic operations on those sets. Also as with the Kripke semantics, this is possible because an overtly modal semantic primitive is appealed to: in the one case *accessible possible world*, in the other case *incompatibility*. (A significant difference is that I have offered a normative, deontic, pragmatic metavocabulary in which to *say* what you have to *do* to deploy that modal semantic primitive, and hence, eventually, the modal operators semantically defined in terms of it.)

The operators defined by the extensional incompatibility semantics are strongly *intensional*, however. We have noticed that one cannot in general compute the incompatibilities of a conjunction from the incompatibilities of its conjuncts. For something can be incompatible with a conjunction without being incompatible with either of its conjuncts. And things are much worse with negation. The two commitments:

- to defining p as incompatibility-entailing q just in case everything incompatible with q is incompatible with p, and
- to understanding the negation of p as its inferentially weakest incompatible, that is, as what is incompatibility-entailed by everything incompatible with p,

together have as a consequence that, to be incompatible with *not-p* is just to be in the intersection of the incompatibility-sets of everything incompatible with p. But that means that we can hold fixed what is incompatible with p, and by varying the incompatibility-sets of some of *those* elements alter the incompatibility-set of *not-p*. It follows that in each

incompatibility-interpretation, the semantic value of *not-p* is *not* determined by the semantic value of *p* alone, but only by it together with the semantic values of a *lot* of other sentences not mentioned in the formula—namely those incompatible with those incompatible with *p*.

It is perhaps less surprising that the incompatibility definitions of what is expressed by necessity and possibility are also intensional, in much the same sense that negation is. So for instance, what is incompatible with *possibly-p* is what is *in*compatible with everything *compatible* with something *compatible* with *p*. Once again, we can fix the semantic interpretant of *p*, its incompatibility set, and still vary the semantic interpretant of *possibly-p*, by varying the semantic interpretants of things compatible with what is compatible with *p*. And the same phenomenon is exhibited by the incompatibility definition of *necessarily-p*.

This is to say that the classical and modal-logical connectives, as semantically defined by incompatibilities, do not have the *semantic sub-formula property*. That is, it is *not* the case that the semantic interpretants of logical compounds formed by applying those connectives is a function of the semantic interpretants of their components, to which the connectives are applied. It is *not* possible to compute the semantic values of arbitrary logical compounds of primitive sentences just from the semantic values of the sentences and the connectives from which they are formed. Another way to put this point is that the incompatibility semantics for these connectives is *not compositional*. It is in this precise sense a *holistic* semantics, in that what is incompatible with (and hence an element of the semantic value of) *not-p* or *necessarily-p* or *possibly-p* depends on what is incompatible with (and hence on the semantic value of) other sentences *q* linked with *p* in that they are compatible or incompatible with it, or incompatible with something that entails it, or compatible with something compatible with it. The holistic character of incompatibility semantics—whether for logical expressions such as 'not' or material, non-logical ones such as 'triangular'—is a result of its codifying the so-to-speak *horizontal* dimension of semantic content, the one that is articulated by the relations of sentences to each other, rather than the *vertical* dimension, which consists in their relations to things that are not themselves sentences.

It is widely believed, and has been particularly forcefully argued by Jerry Fodor, that no holistic semantics can account either for the *projectibility* of language or for its *systematicity*, and hence not for its *learnability*. That

is, it is argued that *only* on the assumption that semantics is *compositional* can we account for the determinateness of the semantic values of an indefinite number of novel compounds of simple expressions, for the fact that wherever some syntactic combinations of those simple expressions have semantic values so do others systematically related to them, and for the fact that speakers of a language can produce and understand an indefinite number of novel compounds, systematically related to one another by their modes of formation, upon mastering the use of the simple expressions and modes of formation.

But I think we are now in a position to see that those arguments cannot be right. They depend upon systematically overlooking the possibility of semantic theories that have the shape of the incompatibility semantics for classical and modal-logical vocabulary we have been considering. For—and this is the key point—although that semantics is *not compositional*, it *is* fully *recursive*. The semantic values of logically compound expressions are wholly determined by the semantic values of logically simpler ones. It is *holistic*, that is, *non*-compositional, in that the semantic value of a compound is *not* computable from the semantic values of its components. But this holism *within* each level of constructional complexity is entirely compatible with recursiveness *between* levels. And this is not just a philosophical *claim* of mine. The system I am describing allows us to *prove* it. (In this context, proof is the word made flesh.)

The semantic values of all the logically compound sentences are computable entirely from the semantic values of *less complex* sentences. It is just that one may need to look at the values of *many*—in the limit *all*—the less complex sentences, not just the ones that appear as sub-formulae of the compound whose semantic value is being computed. The semantics is *projectible* and *systematic*, in that semantic values are determined for all syntactically admissible compounds, of arbitrary degrees of complexity. It is *learnable*—at least in principle, putting issues of contingent psychology aside, in the ideal sense we have been working with. For the capacity to distinguish the incompatibility-sets of primitive propositions is, in the context of the semantic definitions of the connectives in terms of incompatibilities I have offered, *PP-sufficient by algorithmic elaboration* for the capacity to distinguish the incompatibilities of all their logical (including modal-logical) compounds—and hence for the practical capacity to distinguish what is a consequence of what.

What semantic *projectibility, systematicity*, and *learnability-in-principle* require, then, is not semantic *atomism* and *compositionality*, but semantic *recursiveness* with respect to complexity. That is entirely compatible with the semantics being *holistic*, in the sense of *lacking* the *semantic sub-formula property*, which is the hallmark of atomism and compositionality. And the argument for this claim is not merely the description of an abstract possibility. The incompatibility semantics for logical vocabulary provides an up-and-running counterexample to the implicit assumption that semantic recursiveness is achievable only by compositionality. Having compound expressions exhibit the semantic sub-formula property is only *one* way of securing recursiveness. The standard arguments for semantic compositionality are fallacious.[10]

So here is another side-benefit of or philosophical spin-off from looking analytically at *pragmatically mediated* semantic relations between antecedently philosophically interesting vocabularies—and a valuable lesson we can learn from the incompatibility semantics that arose from thinking about complex resultant meaning-use relations between normative and modal vocabularies.

7 Consequence-intrinsic logic

The order of explanation I have been pursuing up to this point,

- starts with practices of giving and asking for reasons,
- argues that they are PP-sufficient for practices of deploying basic *normative* vocabulary—in particular the deontic modal vocabulary of 'commitment' and 'entitlement',
- uses that as a pragmatic metavocabulary that specifies how to deploy a *modal* concept of <u>incompatibility</u>,

---

[10] A more charitable way to put things would be that compositionality—which really amounts to semantic recursiveness—has been confused with the semantic sub-formula property. Thus Jerry Fodor and Ernest Lepore's *The Compositionality Papers* (Oxford University Press, 2002) opens with this definition: ''Compositionality is the property that a system of representations has when (i) it contains both primitive symbols and symbols that are syntactically and semantically complex; and (ii) the latter inherit their syntactic/semantic properties from the former'' (p.1). On this definition, the incompatibility semantics is fully compositional. But Fodor and Lepore go on to draw atomistic consequences from compositionality that in fact only follow from the semantic sub-formula property. So the confusion I am concerned to point out is in play, however we decide to specify it.

- uses that as the basic semantic metavocabulary in which to define a *consequence* relation of incompatibility-entailment,
- and on that basis offers semantic definitions of logical vocabulary, including modal operators.

It is possible to exploit the pragmatic and semantic relations appealed to in this approach in service of a *different*, converse, order of explanation, however. In particular, instead of defining a semantic consequence relation in terms of a *prior* notion of <u>incompatibility</u>, we can *start* with a *consequence* relation—either a *logical* consequence relation or a *material* one that depends on the contents of the *non*-logical vocabulary articulating its premises and conclusions—and *impute* an incompatibility relation on that basis so as semantically to generate just that consequence relation by the procedures I have already put in place.

The idea is to hold fixed the principle that Y is a semantic consequence of X just in case everything incompatible with Y is incompatible with X, but to use that principle relating them to define an incompatibility relation among sets of sentences of the language that would generate whatever consequence relation we are given to begin with. To make this work, we have to ask what conditions a consequence relation defined on an arbitrary set of sentences must meet in order to make it possible to define from it an incompatibility relation such that sets of sentences X and Y stand in the consequence relation (which I'll write '$X \vdash Y$') just in case everything incompatible with Y is incompatible with X (which I will continue to write '$X \models Y$').

It turns out that two conditions suffice:

1. General Transitivity: $\quad \forall X, Y, Z, W \subseteq L[(X \vdash Y \,\&\, Y \cup W \vdash Z) \to X \cup W \vdash Z]$.
2. Defeasibility: $\quad \forall X, Y \subseteq L[\sim(X \vdash Y) \to \exists Z \subseteq L[\forall W \subseteq L[Y \cup Z \vdash W] \,\&\, \exists W \subseteq L[\sim(X \cup Z \vdash W)]]]$.

I will call any consequence relation meeting these conditions 'standard'. The first is a very minimal condition on consequence relations, which corresponds to the usual 'Cut' rule of sequent calculi:

$$\frac{\Gamma : A \text{ and } \Delta, A : B}{\Gamma, \Delta : B}$$

The second says that if Y is *not* a consequence of X, then there is something that yields an absurdity—something that has *everything* as a

consequence—when added to Y but not when added to X. In Appendix 4 to this lecture, I show that if a consequence relation meets these two conditions, then it is possible to define an incompatibility relation that will generate exactly that consequence relation as incompatibility-entailment, by identifying incoherent sets as those that have all sets as their consequences, and then taking two sets to be incompatible if and only if their union is incoherent. That is, for every standard consequence relation, we can find a standard incompatibility relation that semantically validates it.

How interesting the representation theorem that comprises these soundness-and-completeness results is depends on how severe a constraint on consequence relations that second condition is. The easiest way to assess that is to see what familiar, or otherwise interesting, consequence relations do and do not satisfy it. I have already argued, in effect, that sentences attributing ordinary compatible families of incompatible properties—paradigmatically, shapes and colors, and membership in various biological or physical kinds, but encompassing a great many others as well—exhibit material consequence relations that are *standard* in this sense. This is because those sentences stand in the material consequence relations that are defined by their incompatibilities, and those, the representation theorem shows, meet the two conditions of standardness. But what of others, which are not defined to begin with in terms of incompatibility? The consequence relations we understand best are *logical* consequence relations, defined on logical vocabulary by various sets of axioms concerning derivability. Perhaps at this point it comes as no surprise that the consequence relation characteristic of classical logic is a standard one. In that setting, a set of sentences can be taken to be *incoherent* just in case it is *inconsistent*, in that for some p, both p and $\sim p$ can be derived from it. Any superset of an inconsistent set is inconsistent, and from inconsistent sets one can derive everything. Treating two sets as *incompatible* just in case their union is *inconsistent* then yields the classical consequence relation under our usual definition of semantic incompatibility-entailment. And this result holds for the first-order quantificational calculus just as it does for the classical propositional calculus. The consequence relations of both logics are standard, and so can be completely codified semantically by incompatibility relations.

Logically inconsistent sets play just the same role in the consequence relations of standard Lewis modal systems, such as S4 and S5, as they do in

the classical logic on which they are based. So these too are standard, and so semantically codifiable as incompatibility-entailment relations.

Now, the observation with which I shall close is this: our representation theorem shows that any consequence relation that meets the conditions of standardness—whether it be a material or a logical consequence relation—can be codified by a standard incompatibility relation definable in a natural way from that consequence relation. And we have seen that any standard incompatibility relation has a logic whose non-modal vocabulary behaves classically and whose modal vocabulary is S5, in the sense that the natural semantic definitions of such vocabulary in terms of incompatibility yields that logic. Putting these results together, we can say that in this precise sense, S5 (whose non-modal fragment is just classical logic) is the logic *intrinsic to* standard incompatibility relations, and hence standard consequence relations. But since not only classical logic, but *all* the usual modal logics—not only S5, but K, T, S3, S4, and B, have standard consequence relations, classical logic and S5 are the intrinsic logic of, for instance, S4, as well as the others. And although the consequence relation of intuitionistic logic is not standard, so not codifiable by a standard incompatibility relation, in a natural sense it does implicitly contain a standard consequence relation, and so in this somewhat extended sense it, too, has PC + S5 as its intrinsic logic.[11] And in the same sense, so does intuitionistic S4. Relevance logic aside, the logic that is in this sense *intrinsic* to the consequence relations of *most* other familiar logics is classical S5. **S5 accordingly has some claim to being *the* modal logic of consequence relations, whether material or logical**.

The concept of the logic that is *intrinsic* to the consequence relation characteristic of some vocabulary (whether logical or not) is the concept of a new kind of semantic relation between vocabularies. It is mediated by the vocabulary of incompatibility, in terms of which, on the one hand, the consequence relation is codified.

I have been concerned to fill in the three sets of practices that implement the basic VV-sufficiency relations of which the relation of *intrinsicness* of a logic to a vocabulary is the resultant:

- imputing a standard *incompatibility* relation from a standard *consequence* relation (P1 of Figure 5.4);

---

[11] I discuss this point further in Appendix 4.2.3 below.
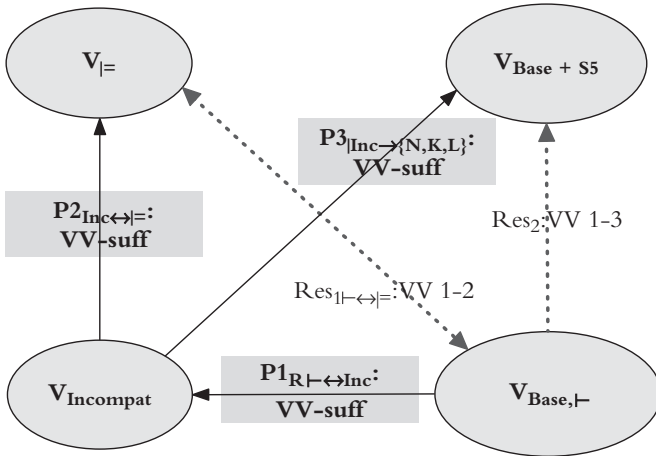
Figure 5.4 Consequence–intrinsic logic

- defining incompatibility-entailment in terms of that incompatibility relation (P2); and
- semantically introducing logical vocabulary, including modal vocabulary, in terms of incompatibility (P3).

So here is another payoff from the metaconceptual apparatus of meaning-use analysis—one that played no part in its initial motivation. The complex resultant MUR this constellation of basic MURs defines is a semantic relation that, apart from this methodology, we would never have been in a position to notice: the relation between logical vocabularies and other vocabularies, when the logical vocabulary is *intrinsic* to the consequence relation *characteristic* of the other vocabulary.

Having put the technical material behind us, in the final lecture I will shift focus by turning attention to what is expressed by *intentional* vocabulary and take up once again, from yet a different perspective, the issue of the relations between normative and modal vocabularies—or if you prefer, between deontic and alethic modalities—as it bears on the nature of intentionality, which itself shows up in this context as itself a *pragmatically mediated semantic relation*.

Technical appendices[12]

Appendix 1: Incompatibility semantics

1 Definitions and axioms

1.1 Incoherence, incompatibility, and entailment   We are given a language L, which is a set of sentences. L may be purely atomic, or it may contain logically complex formulae. L is *proper* if, for each $p \in L$ and each q a subformula of p, we also have $q \in L$. All languages under consideration are proper.

> **Axiom (Persistence)** For finite $X, Y \subseteq L$, and $X \subseteq Y$, if $X \in$ Inc then $Y \in$ Inc.

Incoherence is a generalization of *inconsistency* to the case of non-logical properties. Persistence says that if a set is incoherent, adding more sentences to it cannot cure that condition. An ordered pair $< L, \text{Inc} >$ whose second element is an incoherence property defined over the first element is a *standard incoherence frame* on L. (Henceforth all frames are understood to be standard—that is, their incoherence property satisfies Persistence.)

> **Incompatibility** An *incompatibility function* I is a function from $P(L)$ to $P(P(L))$.

Incoherence properties are related one-to-one to incompatibility functions by:

> **Axiom (Partition)** $X \cup Y \in$ Inc iff $X \in I(Y)$.

---

[12] The formal work presented in these appendices is the result of a collaboration with my research assistant (and Pitt PhD student) Alp Aker. I came up originally with versions of the semantic definitions, the introduction rules for the connectives, proofs of the validity of the various logical principles involving those connectives, and most of the other results reported in these appendices. Alp vastly improved our understanding of the incompatibility semantics by shifting to a definition of incompatibility entailment that is disjunctive on the right (I had used one that was conjunctive on the right). This made it possible for him to formulate the reduction formulae for the connectives, which made all the proofs cleaner and easier. (I had been working directly from the basic definitions, which required extremely laborious derivations from very quantificationally complex formulae.) It also made it possible for him to prove the crucial metatheorems showing that the semantic connective definitions determine extensions of incompatibility frames over a set of non-logical sentences to arbitrarily complex logical compounds of them in a way that is inferentially conservative and unique. (The explicit recursions I had attempted to use to the same end proved unworkable.) Alp is also responsible for the proofs of completeness, soundness, and compactness in Appendix 3.

That is, two sets of sentences are incompatible just in case their union is incoherent. An ordered pair $< L, I >$ is a *standard incompatibility frame* on L. Note that incompatibility is symmetric: if $X \in I(Y)$, then $Y \in I(X)$. It also follows from the Persistence of incoherence that if $X \in I(Y)$ and $X \subseteq X'$, then $X' \subseteq I(Y)$.

Given an incoherence property or an incompatibility function we have the following relation of *incompatibility-entailment*:

Entailment $X \models_I Y$ iff $\bigcap_{p \in Y} I(\{p\}) \subseteq I(X)$.

X can be an infinite set of formulae, but *we require Y to be finite*. When Y is empty we read $\bigcap_{p \in Y} I(\{p\})$ as equivalent to $P(L)$. (Thus $X \models_I \emptyset$ is equivalent to $X \in \text{Inc}$.) We index entailment relations by incompatibility functions (or, equivalently, by incoherence properties). The underlying idea is that one sentence incompatibility–entails another if and only if everything incompatible with the conclusion is incompatible with the premise. That idea is generalized to a relation between sets in a convenient and natural way. The heuristic meaning of $X \models \{y_1, \ldots, y_n\}$ is that X entails y_1 *or ... or* y_n.

Validity X is *valid* iff $Y \in \bigcap_{p \in X} I(\{p\}) \Rightarrow Y \in \text{Inc}$.

Thus $\models \{p\}$ is equivalent to $\{p\}$'s being valid.

1.2 Connective definitions   We have three axioms that govern the behavior of the connectives N, K, and L (which are introduced, and their definitions motivated, in the body of the text):

Axiom (Negation Introduction; NI) $X \cup \{Np\} \in \text{Inc}$ iff $X \models \{p\}$.

Axiom (Conjunction Introduction; CI) $X \cup \{Kpq\} \in \text{Inc}$ iff
$$X \cup \{p, q\} \in \text{Inc}.$$

Axiom (L Introduction; LI) $X \cup \{Lp\} \in \text{Inc}$ iff $X \in \text{Inc}$ or
$$\exists Y[X \cup Y \notin \text{Inc} \ \& \ Y \not\models \{p\}].$$

For ease of reading we sometimes drop brackets around sets and sometimes use the comma to denote set union. Thus we can, for example, write $I(p, X)$ instead of $I(\{p\} \cup X)$ and $X, Np \models q$ instead of $X \cup \{Np\} \models \{q\}$. We also write Apq as an abbreviation of $NKNpNq$, Mp for $NLNp$, and Cpq for $NKpNq$.

2 Basic lemmas

2.1 (Weakening) If $X \models Y$, then $X, W \models Y, V$.

Proof: Suppose $X \models Y$. Then $\bigcap_{p \in Y} I(p) \subseteq I(X)$. Since $I(X) \subseteq I(X, W)$ and $\bigcap_{p \in Y \cup V} I(p) \subseteq \bigcap_{p \in Y} I(p)$, we know $\bigcap_{p \in Y \cup V} I(p) \subseteq I(X, W)$. Thus $X, W \models Y, V$.

2.2 (Cut) If $X \models q, Y$ and $q, W \models V$ then $X, W \models Y, V$.

Proof: We want $\bigcap_{p \in Y \cup V} I(p) \subseteq I(X, W)$. So suppose $S \in \bigcap_{p \in Y \cup V} I(p)$. We then want $S \in I(X, W)$. This is equivalent to $S \cup W \in I(X)$, which—because $X \models q, Y$—holds if both $S \cup W \in I(q)$ and $S \cup W \in I(p)$ for all $p \in Y$.

Now, because $q, W \models V$ and by supposition $S \in I(p)$ for all $p \in V$, we know $S \in I(W, q)$ and thus $S \cup W \in I(q)$. We also know by supposition that $S \in I(p)$ for all $p \in Y$ and so $S \cup W \in I(p)$ for all $p \in Y$.

3 Some modal properties

We begin with two small points:

3.1 $\exists Y[Y \not\models \emptyset \ \& \ Y \not\models p]$ iff $\not\models p$.

Proof: ($\Rightarrow$) Instantiate to get $X \not\models p$. This implies $\not\models p$.

($\Leftarrow$) We show the contrapositive. Suppose $\forall Y(Y \models \emptyset$ or $Y \models p)$. Since $Y \models \emptyset$ implies $Y \models p$ we have $\forall Y(Y \models p)$. Then if X is incompatible with p it is incompatible with everything and so $\models p$.

3.2 $Lp \models \emptyset$ iff $\models \emptyset$ or $\not\models p$.

Proof: Instantiating the L Introduction rule with $\emptyset$ as X we get $Lp \models \emptyset$ iff $\models \emptyset$ or $\exists Y[Y \not\models \emptyset \ \& \ Y \not\models p]$. By 3.1 the latter disjunct is equivalent to $\not\models p$, and we thus have $Lp \models \emptyset$ iff $\models \emptyset$ or $\not\models p$.

Note that the disjunct $\models \emptyset$ ("the True implies the False") is not necessarily false. It is equivalent to $P(L) \subseteq Inc$, which condition is fulfilled in the degenerate case in which every set of sentences is incoherent.

We now get the **basic observation** about modal formulae:

3.3 $X, Lp \models \emptyset \Leftrightarrow X \models \emptyset$ or $Lp \models \emptyset$.

Proof: ($\Rightarrow$) $X, Lp \models \emptyset$ is by definition equivalent to $X \models \emptyset$ or $\exists Y[X, Y \not\models \emptyset$ and $Y \not\models p]$. Since $X, Y \not\models \emptyset$ implies $Y \not\models \emptyset$ we know $\exists Y[X, Y \not\models \emptyset \& Y \not\models p]$ implies $\exists Y[Y \not\models \emptyset$ and $Y \not\models p]$, which, instantiating L Introduction with $X = \emptyset$, implies $Lp \models \emptyset$.

($\Leftarrow$) This follows from Persistence.

It is now easy to prove:

3.4 (Necessitation) $\models p \Rightarrow \models Lp$.

Proof: Suppose $\models p$ and $X, Lp \models \emptyset$. We want X incoherent. By the basic observation either $X \models \emptyset$, as desired, or $Lp \models \emptyset$, in which case either $\models \emptyset$ or $\not\models p$. In the former case every set is incoherent, including X. The latter case contradicts our supposition and so can't occur.

The entailment that corresponds to the T axiom CLpp is also easy:

3.5 $Lp \models p$.

Proof: Suppose not. Then there is some Z such that $Z, p \models \emptyset$ and $Z, Lp \not\models \emptyset$. From the latter it follows that $Z \not\models \emptyset$ and $Lp \not\models \emptyset$. From $Lp \not\models \emptyset$ it follows that $\models p$ by 3.2. But then since $Z, p \models \emptyset$ and p is valid, we know that $Z \models \emptyset$, which implies $Z, Lp \models \emptyset$, which is a contradiction.

Using 3.5 and Cut we easily get the following useful rule:

3.6 $\models Lp \Rightarrow \models p$.

We can also extend the basic observation:

3.7 $X, Lp \models Y$ iff $X \models Y$ or $Lp \models \emptyset$.

Proof: $X, Lp \models Y$ is $\forall Z(Z \in \bigcap_{p \in Y} I(p) \Rightarrow X, Lp, Z \models \emptyset)$, which by 3.3 is $\forall Z(Z \in \bigcap_{p \in Y} I(p) \Rightarrow X, Z \models \emptyset$ or $Lp \models \emptyset)$. This in turn is equivalent to $\forall Z(Z \in \bigcap_{p \in Y} I(p) \Rightarrow X, Z \models \emptyset)$ or $Lp \models \emptyset$, which is $X \models Y$ or $Lp \models \emptyset$.

We can also show that with respect to any particular incompatibility frame, every necessary proposition is either contradictory or valid:

3.8 $Lp \models \emptyset$ or $\models Lp$.

Proof: Suppose $Lp \not\models \emptyset$. Then $\models p$ by 3.2 and so $\models Lp$ by 3.4.

Of course, this does not mean that every necessary proposition is either incoherent in all frames or valid in all frames, but only that it is incoherent-or-valid in all frames.

The next result is a dual of 3.7:

3.9 $X \models Y, Lp$ iff $X \models Y$ or $\models Lp$.

Proof: ($\Rightarrow$) Suppose not. Then $X \models Y$ and $\not\models Lp$. By 3.8 we then know $Lp \models \emptyset$. Since $X \not\models Y$ there is some $Z \in \bigcap_{p \in Y} I(p)$ with $Z \notin I(X)$. Since $Lp \models \emptyset$ we know Z is incompatible with Lp. Since $X \models Y, Lp$ we then have Z incompatible with X, which is a contradiction.

($\Leftarrow$) This follows from Weakening.

4 Semantic reduction

We can, for any given entailment $X \models Y$, show that it is equivalent either to another entailment that mentions fewer connectives or to a Boolean combination of such entailments. We call these equivalences "reduction schemata." They greatly facilitate theorem proving, and their existence makes possible metatheoretical results such as the semantic reduction material in the next section.

4.1 Reduction schemata for non-modal connectives

4.1.1 (Left Negation) $X \models Y, p$ iff $X, Np \models Y$.

Proof: ($\Rightarrow$) Suppose $X \models Y, p$. We want $\bigcap_{r \in Y} I(r) \subseteq I(X, Np)$. Now suppose $Z \in \bigcap_{r \in Y} I(r)$. Then $Z \cup \{Np\} \in \bigcap_{r \in Y} I(r)$. We also know $p \in I(Np)$ and so $Z \cup \{Np\} \in I(p)$. Then since $X \models p, Y$ we have $Z \cup \{Np\} \in I(X)$ and so $Z \in I(X, Np)$.

($\Leftarrow$) Suppose $X, Np \models Y$. We want $\bigcap_{r \in Y \cup \{p\}} I(r) \subseteq I(X)$. So suppose $Z \in \bigcap_{r \in Y \cup \{p\}} I(r)$. Since $X, Np \models Y$ we have $Z \in I(X, Np)$. Then $Z, X \models p$. Since $Z \in I(p)$ Cut gives $Z \in I(Z, X)$ and thus $Z \in I(X)$.

4.1.2 (Right Negation) $X \models Y, Np$ iff $X, p \models Y$.

Proof: ($\Rightarrow$) Suppose $X \models Y, \mathrm{N}p$. We want $\bigcap_{r \in Y} \mathrm{I}(r) \subseteq \mathrm{I}(X, p)$. So suppose $Z \in \bigcap_{r \in Y} \mathrm{I}(r)$. Then $Z \cup \{p\} \in \bigcap_{r \in Y} \mathrm{I}(r)$. We also know $Z \cup \{p\} \in \mathrm{I}(\mathrm{N}p)$. Since $X \models Y, \mathrm{N}p$ it follows that $Z \cup \{p\} \in \mathrm{I}(X)$, or $Z \in \mathrm{I}(X, p)$.

($\Leftarrow$) Suppose $X, p \models Y$. We want $\bigcap_{r \in Y \cup \{\mathrm{N}p\}} \mathrm{I}(r) \subseteq \mathrm{I}(X)$. So suppose $Z \in \bigcap_{r \in Y \cup \{\mathrm{N}p\}} \mathrm{I}(r)$. Then since $Z \in \bigcap_{r \in Y} \mathrm{I}(r)$ and $X, p \models Y$ we have $Z \in \mathrm{I}(X, p)$ or $Z, X, p \models \emptyset$. Since $Z \in \mathrm{I}(\mathrm{N}p)$ we have $Z \models p$. From $Z, X, p \models \emptyset$ and $Z \models p$ we can apply Cut to get $Z, X \models \emptyset$, or $Z \in \mathrm{I}(X)$, as desired.

4.1.3 (Left Conjunction) $X, \mathrm{K}pq \models Y$ iff $X, p, q \models Y$.

Proof: By definition $\mathrm{I}(X, \mathrm{K}pq) = \mathrm{I}(X, p, q)$. The result follows immediately.

4.1.4 (Right Conjunction) $X \models Y, \mathrm{K}pq$ iff $X \models Y, p$ and $X \models Y, q$.

Proof: ($\Rightarrow$) Suppose $X \models Y, \mathrm{K}pq$. Then $\bigcap_{r \in Y \cup \{\mathrm{K}pq\}} \mathrm{I}(r) \subseteq \mathrm{I}(X)$. Equivalently, if $Z \in \mathrm{I}(p, q)$ and $Z \in \bigcap_{r \in Y} \mathrm{I}(r)$ then $Z \in \mathrm{I}(X)$. But $\mathrm{I}(p) \subseteq \mathrm{I}(p, q)$. Then if $Z \in \mathrm{I}(p)$ and $Z \in \bigcap_{r \in Y} \mathrm{I}(r)$ it follows that $Z \in \mathrm{I}(p, q)$ and $Z \in \bigcap_{r \in Y} \mathrm{I}(r)$ and so $Z \in \mathrm{I}(X)$. Thus $X \models Y, p$. We can argue similarly to get $X \models Y, q$.

($\Leftarrow$) Suppose $X \models Y, p$ and $X \models Y, q$. We want $\bigcap_{r \in Y \cup \{\mathrm{K}pq\}} \mathrm{I}(r) \subseteq \mathrm{I}(X)$. So suppose $Z \in \mathrm{I}(\mathrm{K}pq)$ and $Z \in \bigcap_{r \in Y} \mathrm{I}(r)$. If $Z \in \mathrm{I}(\mathrm{K}pq)$ then $Z, p, q \models \emptyset$. By Cut and the fact that $X \models Y, p$ we then have $Z, X, q \models Y$. Applying Cut again, this time with $X \models Y, q$, we get $Z, X \models Y$. Since by supposition $Z \in \bigcap_{r \in Y} \mathrm{I}(r)$ we then know $Z \in \mathrm{I}(Z, X)$, which is $Z \in \mathrm{I}(X)$, as desired.

4.2 Reduction schemata for modal connectives

4.2.1 (Left Necessity) $X, \mathrm{L}p \models Y$ iff $X \models Y$ or $\not\models p$.

Proof: If we are in the degenerate frame everything implies everything else, so the result holds. If not, it follows from 3.7 and 3.2.

4.2.2 (Right Necessity) $X \models Y, \mathrm{L}p$ iff $X \models Y$ or $\models p$.

Proof: Apply 3.9, then 3.4 and 3.6.

5 Incompatibilities for extensions of a language

5.1 Motivation   Crucial to the compositionality of meaning is that the semantic values of logically complex sentences be reducible to the semantic values of their constituents. In the framework of incompatibility logic, however, meaning is holistic, and so this reduction cannot proceed sentence by sentence. What we want instead is to be show how the frame for a language with logically complex sentences can be reduced to the frame for a syntactically less complex fragment of the language.

Suppose $L \subseteq L'$. Say that L' is a *proper extension* of L if L' is a proper language in the sense of section 1 and all atomic formulae in L' are contained in L (thus every formula in L' is built from formulae in L and all the intermediate formulae in the construction are also in L'). Our problem, then, can be formulated thus: Given a frame Inc for a language L, and given a language L' properly extending L, what is the frame Inc' for L' that is determined by Inc?

One property that Inc' should have is that it should agree with Inc about the semantic properties of vocabulary in L. For example, if L is an atomic language and L' contains in addition logically complex formulae formed from the sentences of L, Inc' should agree with Inc about which entailments $X \models Y$ hold for those X, Y that are sets of atomic formulae. This agreement is a form of *inferential conservativeness*:

Let $L \subseteq L'$ and let Inc be a frame for L. Then a frame Inc' for L' is *inferentially conservative (for short, IC) with respect to Inc* if, for $X, Y \subseteq L$, $X \models_{\text{Inc}} Y \Leftrightarrow X \models_{\text{Inc}'} Y$.

For finite languages, the requirement of inferential conservativeness is sufficient to determine frames for properly extending languages. That is, given a finite L, a frame Inc for L, and an extending language L', it is sufficient to determine Inc' simply to require that Inc' be inferentially conservative with respect to Inc.

If L is infinite, however, it ceases to be the case that atomic frames generate unique inferentially conservative extensions. (Inferential conservativeness suffices for infinite languages as well if we stipulate that all incoherent sets be finite. But there is no compelling a priori reason for such a stipulation.) Given a frame Inc for some language L, and a language L' extending L, there can be multiple frames for L' that conservatively extend Inc. But as algebraic experience might lead us to expect, there is always a

single *smallest* frame for L' that is inferentially conservative with respect to Inc (where 'smallest' has the sense of, contained in every other frame for L' that is IC with respect to Inc). There might, depending on the case, be *other* frames for L' that are inferentially conservative with respect to Inc, but each of these other frames properly contains Inc', i.e., they can be obtained from Inc' by *adding* further semantic information to Inc' in the form of stipulating that additional sets of formulae are incoherent in addition to those deemed incoherent by Inc'. Accordingly, we view Inc', but not these other frames, as determined solely by the semantic information contained in Inc.

Thus, we have the following definition that encapsulates all that we need for a sensible theory of semantic reduction:

Let L' be a proper extension of L and let Inc be a frame for L. The *frame for L' determined by Inc* is the smallest frame for L' that is IC with respect to Inc.

5.2 Existence of the determined frame   We now show that the determined frame exists. (If it does exist, it is immediate from the definition that it is unique.) The following results can be heuristically summarized as follows. Suppose we are given a language L, a frame Inc for L, and a proper extension L' of L. Now suppose a frame for L' contains only subsets of L' whose incoherence can be shown to follow from Inc by a finite number of applications of the reduction schemata of section 4. Such a frame is contained in every frame for L' that is IC with respect to Inc, and hence it is the unique smallest frame for L' that is IC with respect to Inc (5.2.1−5.2.2). Further, such a frame for L' always exists (5.2.3−5.2.7). Thus the frame for L' determined by Inc exists (5.2.8).

Call a frame Inc *finitary over L* if for each $X \in$ Inc, $X - L$ is finite (that is, X contains only finitely many formulae not in L). Call a frame Inc *genetically finitary over L* if for each $X \in$ Inc there is some $X' \subseteq X$ such that $X' \in$ Inc and $X' - L$ is finite.

5.2.1 (Semantic Reduction Lemma) Given a language L and frame Inc, let $L_\emptyset$ be a fragment of L such that L is a proper extension of $L_\emptyset$. Let $X, Y \subseteq L$ be such that $X - L_\emptyset$ and $Y - L_\emptyset$ are finite. Then there are a Boolean function F on n propositions and sets of sentences $X_i, Y_i \subseteq L_\emptyset$ such that $X \models_{\text{Inc}} Y$ iff $F(X_1 \models_{\text{Inc}} Y_1$;

$\ldots; X_n \models_{\mathrm{Inc}} Y_n$). Further, $F(X_1 \models_{\mathrm{Inc}} Y_1; \ldots; X_n \models_{\mathrm{Inc}} Y_n)$ can be chosen on the basis of just the syntax of the members of X and Y.

Proof: We actually show a slightly stronger result, viz. $F(X_1 \models_{\mathrm{Inc}} Y_1; \ldots; X_n \models_{\mathrm{Inc}} Y_n)$ iff $G(X'_1 \models_{\mathrm{Inc}} Y'_1; \ldots; X'_m \models_{\mathrm{Inc}} Y'_m)$, where $X_i, Y_i \subseteq L$; $X'_i, Y'_i \subseteq L_\emptyset$; and $X_i - L_\emptyset$ and $Y_i - L_\emptyset$ are finite for every i.

Fix an ordering of the formulae of L. We prove the result by induction on the number of connectives contained in all the X_i, Y_i.

Given $F(X_1 \models_{\mathrm{Inc}} Y_1; \ldots; X_n \models_{\mathrm{Inc}} Y_n)$ choose the first $X_i \models_{\mathrm{Inc}} Y_i$ that mentions a formula not in $L_\emptyset$ and choose the first such formula in $X_i \models_{\mathrm{Inc}} Y_i$ according to our chosen ordering. We have six possibilities, according as the major operator is N, K, or L, and according as the formula is in the antecedent or the consequent of the entailment. In the first case (an N-formula in antecedent position), we have

$$F(X_1 \models_{\mathrm{Inc}} Y_1; \ldots; X_n \models_{\mathrm{Inc}} Y_n) \text{ iff}$$
$$F(X_1 \models_{\mathrm{Inc}} Y_1; \ldots; \mathrm{N}p, Z_i \models_{\mathrm{Inc}} Y_i; \ldots; X_n \models_{\mathrm{Inc}} Y_n),$$

where $X_i = \{\mathrm{N}p\} \cup Z_i$. Applying 4.1.2 we get:

$$F(X_1 \models_{\mathrm{Inc}} Y_1; \ldots; X_n \models_{\mathrm{Inc}} Y_n) \text{ iff}$$
$$F(X_1 \models_{\mathrm{Inc}} Y_1; \ldots; Z_i \models_{\mathrm{Inc}} Y_i, p; \ldots; X_n \models_{\mathrm{Inc}} Y_n).$$

The right-hand side is a Boolean combination of entailments that mentions one fewer connective than the left-hand side, so we may apply the induction hypothesis to get:

$$F(X_1 \models_{\mathrm{Inc}} Y_1; \ldots; X_n \models_{\mathrm{Inc}} Y_n) \text{ iff } G(X'_1 \models_{\mathrm{Inc}} Y'_1; \ldots;$$
$$X'_m \models_{\mathrm{Inc}} Y'_m)$$

as desired. The remaining five cases are treated similarly, applying 4.1.1, 4.1.3, 4.1.4., 4.2.1, or 4.2.2 as the case may be.

5.2.2 Let Inc be a frame for L and let L' properly extend L. Let Inc$'$ be a frame for L' that is genetically finitary over L and IC with respect

to Inc. If Inc″ is another frame for L' that is IC with respect to Inc, then Inc′ $\subseteq$ Inc″.

Proof: Suppose $X \in$ Inc′. Since Inc′ is genetically finitary over L, there is some $X' \subseteq X$ such that $X' \in$ Inc′ and $X' - L$ is finite. Then by 5.2.1 there exists some F and some $X_i, Y_i \subseteq L$ such that $X' \models_{\text{Inc}'} \emptyset$ implies $F(X_1 \models_{\text{Inc}'} Y_1; \ldots; X_n \models_{\text{Inc}'} Y_n)$. Since Inc′ is IC with respect to Inc, this implies $F(X_1 \models_{\text{Inc}} Y_1; \ldots; X_n \models_{\text{Inc}} Y_n)$. Since Inc″ is IC with respect to Inc, we have $F(X_1 \models_{\text{Inc}''} Y_1; \ldots; X_n \models_{\text{Inc}''} Y_n)$. By 5.2.1 again $X' \models_{\text{Inc}''} \emptyset$, or $X' \in$ Inc″. By Persistence, $X \in$ Inc″. Thus Inc′ $\subseteq$ Inc″.

5.2.3 Let Inc be a frame for L. Let $L' = L \cup \{\text{K}pq\}$ for some $p, q \in L$. Define $F(X)$ as follows. If $X = X' \cup \{\text{K}pq\}$ then $F(X) = X' \cup \{p, q\}$; otherwise $F(X) = X$. Let $X \in$ Inc′ if $F(X) \in$ Inc. Then (i) Inc′ is a frame for L'; (ii) Inc′ is IC with respect to Inc; and (iii) Inc′ is finitary over L.

Proof: (i) We verify the frame axioms.

(Persistence) Suppose $X \in$ Inc′ and $X \subseteq Y$. Then $F(X) \in$ Inc and so by Persistence for Inc, $F(Y) \in$ Inc. Then $Y \in$ Inc′.

(NI) Suppose $X \cup \{\text{N}r\} \in$ Inc′. We want $X \models_{\text{Inc}'} r$. Suppose then that $Z \cup \{r\} \in$ Inc′. It now suffices to show that $Z \cup X \in$ Inc′. Since $X \cup \{\text{N}r\} \in$ Inc′ we know $F(X) \cup \{\text{N}r\} \in$ Inc. Then $F(X) \models_{\text{Inc}} r$. Since $Z \cup \{r\} \in$ Inc′ we know $F(Z) \cup \{r\} \in$ Inc. Then $F(X) \cup F(Z) = F(X \cup Z) \in$ Inc, and so $X \cup Z \in$ Inc′.

Suppose $X \models_{\text{Inc}'} r$. We want $X \cup \{\text{N}r\} \in$ Inc′. Since $\text{N}r \neq \text{K}qp$, $\text{N}r \in L$. Then $\{r, \text{N}r\} \in$ Inc, so $\{r, \text{N}r\} \in$ Inc′. Since $X \models_{\text{Inc}'} r$, it follows that $X \cup \{\text{N}r\} \in$ Inc′.

(KI) $X \cup \text{K}rs \in$ Inc′ iff $F(X \cup \text{K}rs) \in$ Inc
     iff $F(X) \cup \{r,s\} \in$ Inc
     iff $X \cup \{r, s\} \in$ Inc′

(LI) Suppose $X \cup \{\text{L}r\} \in$ Inc′. We want $X \in$ Inc′ or $\exists Y[X \cup Y \notin$ Inc′ & $Y \not\models_{\text{Inc}'} r]$.

Suppose $X \notin$ Inc′. We show $\exists Y[X \cup Y \notin$ Inc′ & $Y \not\models_{\text{Inc}'} r]$. Since $X \cup \{\text{L}r\} \in$ Inc′ we know $F(X) \cup \{\text{L}r\} \in$ Inc. Since $X \notin$ Inc′ we know $F(X) \notin$ Inc. Then $\exists Y[F(X) \cup Y \notin$ Inc & $Y \not\models_{\text{Inc}} r]$.

Instantiate to some W, so that $F(X) \cup W \notin \mathrm{Inc}$ and $W \not\models_{\mathrm{Inc}} r$. Since $\mathrm{K}pq \notin L$ we know $W = F(W)$, so $F(X \cup W) \notin \mathrm{Inc}$ and $F(W) \not\models_{\mathrm{Inc}} r$. Then there is some Z such that $Z \cup \{r\} \in \mathrm{Inc}$ and $Z \cup F(W) \notin \mathrm{Inc}$. Note that $\mathrm{K}pq \notin Z$ because $\mathrm{K}pq \notin L$ and $Z \subseteq L$. Then $Z \cup \{r\} \in \mathrm{Inc}'$ and $Z \cup W \notin \mathrm{Inc}'$. Thus $W \not\models_{\mathrm{Inc}'} r$. Thus $\exists Y[X \cup Y \notin \mathrm{Inc}' \;\&\; Y \not\models_{\mathrm{Inc}'} r]$.

Now suppose $X \in \mathrm{Inc}'$ or $\exists Y[X \cup Y \notin \mathrm{Inc}' \;\&\; Y \not\models_{\mathrm{Inc}'} r]$. We want $X \cup \{\mathrm{L}p\} \in \mathrm{Inc}'$. If $X \in \mathrm{Inc}'$ then the result follows by Persistence. So suppose $\exists Y[X \cup Y \notin \mathrm{Inc}' \;\&\; Y \not\models_{\mathrm{Inc}'} r]$. Instantiate to some W to get $X \cup W \notin \mathrm{Inc}'$ and $W \not\models_{\mathrm{Inc}'} r$. Then $F(X \cup W) = F(X) \cup F(W) \notin \mathrm{Inc}'$. Now since $W \not\models_{\mathrm{Inc}'} r$ there is some Z such that $Z \cup \{r\} \in \mathrm{Inc}'$ but $Z \cup W \notin \mathrm{Inc}'$. Then $F(Z) \cup \{r\} \in \mathrm{Inc}$ and $F(Z \cup W) = F(Z) \cup F(W) \notin \mathrm{Inc}$. Thus $F(W) \not\models_{\mathrm{Inc}} r$. It follows that $\exists Y[F(X) \cup Y \notin \mathrm{Inc} \;\&\; Y \not\models_{\mathrm{Inc}} r]$ and so $F(X) \cup \{\mathrm{L}p\} = F(X \cup \{\mathrm{L}p\}) \in \mathrm{Inc}$. Thus $X \cup \{\mathrm{L}p\} \in \mathrm{Inc}'$, as desired.

(ii) Suppose $X, Y \subseteq L$. We want $X \models_{\mathrm{Inc}'} Y \Leftrightarrow X \models_{\mathrm{Inc}} Y$.

($\Rightarrow$) Suppose $X \models_{\mathrm{Inc}'} Y$ and $Z \cup \{y_i\} \in \mathrm{Inc}$ for each $y_i \in Y$. We want $Z \cup X \in \mathrm{Inc}$. Since $Y, Z \subseteq L$ we know $F(Z \cup \{y_i\}) = Z \cup \{y_i\}$. Then $F(Z \cup \{y_i\}) \in \mathrm{Inc}$ and so $Z \cup \{y_i\} \in \mathrm{Inc}'$. Since $X \models_{\mathrm{Inc}'} Y$, $Z \cup X \in \mathrm{Inc}'$. By construction, $F(Z \cup X) = Z \cup X \in \mathrm{Inc}$.

($\Leftarrow$) Suppose $X \models_{\mathrm{Inc}} Y$ and $Z \cup \{y_i\} \in \mathrm{Inc}'$ for each $y_i \in Y$. Then $F(Z \cup \{y_i\}) \in \mathrm{Inc}$, and since $Y \subseteq L$, $F(Z) \cup \{y_i\} \in \mathrm{Inc}$. Since $X \models_{\mathrm{Inc}} Y$, $F(Z) \cup X = F(Z \cup X) \in \mathrm{Inc}$. But then $Z \cup X \in \mathrm{Inc}'$.

(iii) Consider some $X \in \mathrm{Inc}'$. The only formula not in L that X can contain is $\mathrm{K}pq$, so $X - L$ is finite.

5.2.4 Let Inc be a frame for L. Let $L' = L \cup \{\mathrm{N}p\}$ for some $p \in L$. Then let $X \in \mathrm{Inc}'$ if $X \in \mathrm{Inc}$ and let $X \cup \{\mathrm{N}p\} \in \mathrm{Inc}$ if $X \models_{\mathrm{Inc}} p$. Then (i) Inc' is a frame for L'; (ii) Inc is IC with respect to Inc; and (iii) Inc' is finitary over L.

Proof: As in 5.2.3.

5.2.5 Let Inc be a frame for L. Let $L' = L \cup \{\mathrm{L}p\}$ for some $p \in L$. Then let $X \in \mathrm{Inc}'$ if $X \in \mathrm{Inc}$ and let $X \cup \{\mathrm{L}p\}$ if $\not\models_{\mathrm{Inc}} p$. Then (i) Inc' is

a frame for L'; (ii) Inc is IC with respect to Inc; and (iii) Inc′ is finitary over L.

Proof: As in 5.2.3.

5.2.6 Let Inc be a frame for L. Let $< p_1, p_2, ... >$ be a sequence of formulae such that each $L_i = L \cup \{p_1, ..., p_i\}$ is a proper extension of L; let $L_0 = L$. Then (i) there exists a frame Inc_i for each L_i; (ii) Inc_i is IC with respect to Inc_j for each $j \leq i$; and (iii) Inc_i is finitary over L_j for each $j \leq i$.

Proof: By induction on the L_i.

(Base case) L_0 satisfies (i)-(iii) immediately.

(Inductive case) Assume the claim is true for $L_0, ..., L_{n-1}$, and that we thus have for each $i \leq n - 1$ a frame Inc_i for L_i satisfying (ii) and (iii). Now, p_n is either Kpq, Np, or Lp, for some p (and possibly q) in L_{n-1}. Then by 5.2.3, 5.2.4, or 5.2.5, as the case may be, we have a frame Inc_n for L_n. Hence (i) is satisfied. We also know that Inc_n is IC with respect to Inc_{n-1} and that it is finitary over L_{n-1}. We now show that Inc_n satisfies (ii) and (iii).

(ii) Consider some $X, Y \subseteq L_j$ for $j \leq n$. Since $L_j \subseteq L_{n-1}$, we have $X, Y \subseteq L_{n-1}$. And since Inc_n is IC with respect to Inc_{n-1}, we have $X \models_{Inc_n} Y$ iff $X \models_{Inc_{n-1}} Y$. But Inc_{n-1} is IC with respect to Inc_j, so $X \models_{Inc_{n-1}} Y$ iff $X \models_{Inc_j} Y$. Thus $X \models_{Inc_n} Y$ iff $X \models_{Inc_j} Y$, which is to say that Inc_n is IC with respect to Inc_j.

(iii) Consider some $X \in Inc_n$. For $i \leq n$ the formulae in X that are not in L_i are a subset of $L_n - L_i$, which is finite. Thus X is finitary over L_i.

5.2.7 Let Inc be a frame for L. Let $< p_1, p_2, ... >$ be a sequence of formulae such that each $L_i = L \cup \{p_1, ..., p_i\}$ is a proper extension of L. Let $L' = L \cup \{p_1, p_2, ...\}$. Then (i) there exists a frame Inc′ for L'; (ii) Inc′ is IC with respect to Inc; and (iii) Inc′ is genetically finitary over L.

Proof: Let $L_0 = L$. By 5.2.6, there is a frame Inc_i for each L_i; Inc_i is IC with respect to Inc_j for all $j \leq i$; and Inc_i is finitary over L_j. Let $X \in Inc'$ if there is some $X' \subseteq X$ such that $X' \in Inc_i$ for some i. We claim that Inc′ is the desired frame.

Lemma: Let $X \cup \{p\} \in \text{Inc}'$. Then there is some $X' \subseteq X$ and some i such that $X' \cup \{p\} \in \text{Inc}_i$.

Proof: If $X \cup \{p\} \in \text{Inc}'$ then there is some $W \subseteq X \cup \{p\}$ and some j such that $W \in \text{Inc}_j$. Then there is some k, with $j \leq k$, such that $p \in L_k$. Since Inc_k is IC with respect to Inc_j, $W \in \text{Inc}_k$ and so $W \cup \{p\} \in \text{Inc}_k$ by Persistence. Take $X' = W - \{p\}$. QED (lemma).

In the following we use the lemma without notice.

(i) We verify the frame axioms.

(Persistence)   Suppose $X \in \text{Inc}'$ and $X \subseteq Y$. Then there is some $X' \subseteq X$ such that $X \in \text{Inc}_i$ for some i. But $X' \subseteq Y$ and so $Y \in \text{Inc}'$ by construction.

(NI) Suppose $X \cup \{Np\} \in \text{Inc}'$. We want $X \models_{\text{Inc}'} p$. Suppose $Z \cup \{p\} \in \text{Inc}'$. It suffices then to show that $Z \cup X \in \text{Inc}'$. Since $X \cup \{Np\} \in \text{Inc}'$ there is some $X' \subseteq X$ and some j such that $X' \cup \{Np\} \in \text{Inc}_j$. Similarly, there is some k and some $Z' \subseteq Z$ such that $Z' \cup \{p\} \in \text{Inc}_k$. Let $l = \max\{k, j\}$. Then $X' \cup \{Np\} \in \text{Inc}_l$ and $Z' \cup \{p\} \in \text{Inc}_l$. Since $X' \cup \{Np\} \in \text{Inc}_l$ we know $X' \models_{\text{Inc}_l} p$. Then $X' \cup Z' \in \text{Inc}_l$ and so $X \cup Z \in \text{Inc}'$.

Suppose $X \models_{\text{Inc}'} p$. We want $X \cup \{Np\} \in \text{Inc}'$. Choose some i such that $Np \in L_i$. Then $\{p, Np\} \in \text{Inc}_i$ and so $\{p, Np\} \in \text{Inc}'$. Then $X \cup \{Np\} \in \text{Inc}'$.

(KI) Suppose $X \cup \{Kpq\} \in \text{Inc}'$. We want $X \cup \{p, q\} \in \text{Inc}'$. Since $X \cup \{Kpq\} \in \text{Inc}'$ there is some $X' \subseteq X$ and some i such that $X' \cup \{Kpq\} \in \text{Inc}_i$. Then $X' \cup \{p, q\} \in \text{Inc}_i$ and so $X \cup \{p, q\} \in \text{Inc}'$.

The converse is proved similarly.

(LI) Suppose $X \cup \{Lp\} \in \text{Inc}'$. We want $X \in \text{Inc}'$ or $\exists Y[X \cup Y \notin \text{Inc}' \ \& \ Y \not\models_{\text{Inc}'} p]$.

Suppose $X \notin \text{Inc}'$. We show $\exists Y[X \cup Y \notin \text{Inc}' \ \& \ Y \not\models_{\text{Inc}'} p]$. Since $X \cup \{Lp\} \in \text{Inc}'$ there is some $X' \subseteq X$ and some j such that $X' \cup \{Lp\} \in \text{Inc}_j$. Then $X' \in \text{Inc}_j$ or $\not\models_{\text{Inc}_j} p$. Since $X \notin \text{Inc}'$, we know $X' \notin \text{Inc}_j$; thus $\not\models_{\text{Inc}_j} p$. Then there is some Z such that $Z \cup \{p\} \in \text{Inc}_j$ but $Z \notin \text{Inc}_j$. Thus $Z \cup \{p\} \in \text{Inc}'$. Further,

we claim $Z \notin \mathrm{Inc}'$. For suppose $Z \in \mathrm{Inc}'$. Then there is some $Z' \subseteq Z$ and some k such that $Z' \in \mathrm{Inc}_k$. Let $l = \max\{j, k\}$. Then $Z' \in \mathrm{Inc}_l$ because Inc_l is IC with respect to Inc_k. But since $L_j \subseteq L_l$ we know $Z \subseteq L_l$, so $Z \in \mathrm{Inc}_l$ by Persistence. But since $Z \subseteq L_j$ and Inc_l is IC with respect to Inc_j, we have $Z \in \mathrm{Inc}_j$, which is a contradiction. Thus $X \notin \mathrm{Inc}'$ and $\not\models_{\mathrm{Inc}'} p$. Generalizing with $Y = \emptyset$ we have $\exists Y[X \cup Y \notin \mathrm{Inc}' \ \& \ Y \not\models_{\mathrm{Inc}'} p]$, as desired.

Now suppose $X \in \mathrm{Inc}'$ or $\exists Y[X \cup Y \notin \mathrm{Inc}' \ \& \ Y \not\models_{\mathrm{Inc}'} p]$. We want $X \cup \{\mathrm{L}p\} \in \mathrm{Inc}'$.

If $X \in \mathrm{Inc}'$ the result follows by Persistence. So suppose $X \notin \mathrm{Inc}'$ and $\exists Y[X \cup Y \notin \mathrm{Inc}' \ \& \ Y \not\models_{\mathrm{Inc}'} p]$. Instantiate to some W, to give $X \cup W \notin \mathrm{Inc}'$ and $W \not\models_{\mathrm{Inc}'} p$. Then there is some Z such that $Z \cup \{p\} \in \mathrm{Inc}'$ and $Z \cup W \notin \mathrm{Inc}'$. Since $Z \cup W \notin \mathrm{Inc}'$ we know $Z \notin \mathrm{Inc}'$. Since $Z \cup \{p\} \in \mathrm{Inc}'$ there is some $Z' \subseteq Z$ and some i such that $Z' \cup \{p\} \in \mathrm{Inc}_i$. Also, since $Z \notin \mathrm{Inc}'$ we know $Z' \notin \mathrm{Inc}_i$. Then $\not\models_{\mathrm{Inc}_i} p$. Pick some subset $X' \subseteq X$ such that $X' \subseteq L_i$. Since $X \notin \mathrm{Inc}'$ we know $X' \notin \mathrm{Inc}_i$. Then we have $X' \notin \mathrm{Inc}_i$ and $\not\models_{\mathrm{Inc}_i} p$. Generalizing with $Y = \emptyset$ we have $\exists Y[X' \cup Y \notin \mathrm{Inc}_i \ \& \ Y \not\models_{\mathrm{Inc}_i} p]$. Then $X' \cup \{\mathrm{L}p\} \in \mathrm{Inc}_i$ and so $X \cup \{\mathrm{L}p\} \in \mathrm{Inc}'$ by construction.

(ii) Suppose $X, Y \subseteq L$. We want $X \models_{\mathrm{Inc}'} Y \Leftrightarrow X \models_{\mathrm{Inc}} Y$.

($\Rightarrow$) Suppose $X \models_{\mathrm{Inc}'} Y$ and $Z \cup \{y_i\} \in \mathrm{Inc}$ for each $y_i \in Y$. We want $Z \cup X \in \mathrm{Inc}$. Since $Z \cup \{y_i\} \in \mathrm{Inc} = \mathrm{Inc}_0$ we have $Z \cup \{y_i\} \in \mathrm{Inc}'$ and so $Z \cup X \in \mathrm{Inc}'$. Then there is some $W \subseteq Z \cup X$ and some j such that $W \in \mathrm{Inc}_j$. But Inc_j is IC with respect to Inc and $W \subseteq L$, so $W \in \mathrm{Inc}$. Since $W \subseteq Z \cup X \subseteq L$, it follows by Persistence that $Z \cup X \in \mathrm{Inc}$.

($\Leftarrow$) Suppose $X \models_{\mathrm{Inc}} Y$ and $Z \cup \{y_i\} \in \mathrm{Inc}'$ for each $y_i \in Y$. Then for each y_i there is some $Z'_i \subseteq Z$ and some j such that $Z'_i \cup \{y_i\} \in \mathrm{Inc}_j$. Since there are a finite number of y_i (recall the definition of entailment from 1.1), there is a largest such j; call it k. Since Inc_k is IC with respect to Inc_j for each j, we have $Z'_i \cup \{y_i\} \in \mathrm{Inc}_k$ for all i. Let $Z' = \bigcup_i Z'_i$. We know $Z' \subseteq L_k$ and so by Persistence $Z' \cup \{y_i\} \in \mathrm{Inc}_k$. Since Inc_k is IC with respect

to Inc, we know that $X \models_{\mathrm{Inc}_k} Y$. Then $X \cup Z' \in \mathrm{Inc}_k$. Thus $X \cup Z \in \mathrm{Inc}'$, as desired.

(iii) Suppose $X \in \mathrm{Inc}'$. Then there is some $X' \subseteq X$ and some i such that $X' \in \mathrm{Inc}_i$. But Inc_i is finitary over L, so $X' - L$ is finite. In addition $X' \in \mathrm{Inc}'$ by construction.

5.2.8 Let Inc be a frame for L and let L' properly extend L. Then there exists a frame Inc' that is the frame for L' determined by Inc.

Proof: Let $< p_1, p_2, \ldots >$ be an enumeration of $L' - L$ such that $i \leq j$ implies that p_j is not a subformula of p_i; if $L' - L$ is finite then let the sequence repeat after some n. Since L' is a proper extension of L we know $L \cup \{p_1, \ldots, p_i\}$ is a proper language for every i. Then by 5.2.7 there is a frame Inc' for L' that is IC with respect to Inc and is genetically finitary over L. By 5.2.2 Inc' is contained in every frame Inc'' for L' that is IC with respect to Inc. Inc' is therefore the smallest frame for L' that is IC with respect to Inc, and hence is the frame for L' determined by Inc.

Appendix 2: Logic using the reduction formulae

Appendix 1 showed that the semantic definitions of the principal connectives in terms of incompatibility/incoherence underwrites these six reduction schemata:

LN: $X, \mathrm{N}p \models Y \Leftrightarrow X \models Y, p$.
RN: $X \models Y, \mathrm{N}p \Leftrightarrow X, p \models Y$.

LK: $X, \mathrm{K}pq \models Y \Leftrightarrow X, p, q \models Y$.
RK: $X \models Y, \mathrm{K}pq \Leftrightarrow X \models Y, p$ and $X \models Y, q$.

LL: $X, \mathrm{L}p \models Y \Leftrightarrow X \models Y$ or $\not\models p$.
RL: $X \models Y, \mathrm{L}p \Leftrightarrow X \models Y$ or $\models p$.

Using them, it is easy to show that the incompatibility semantics validates classical logic for N and K (hence for A), and S5 when we add L (and hence M):

1 Negation

1.1 If $\{p\} \in I(X)$ and $\{\mathrm{N}p\} \in I(X)$, then $X \in \mathrm{Inc}$.

Proof: If $\{Np\} \in I(X)$ then $X \models p$, that is, $I(p) \subseteq I(X)$. Since $\{p\} \in I(X)$, we have $X \in I(p)$. Thus $X \in I(X)$, or $X \in Inc$.

It follows that $Np \in I(p)$:

1.2 (Double Negation Equivalence) $NNp \approx p$.

Proof: $NNp \models p \Leftrightarrow \emptyset \models Np, p$ (LN) $\Leftrightarrow p \models p$ (RN).
$(\Leftarrow)\ p \models NNp \Leftrightarrow p, Np \models \emptyset$ (RN) $\Leftrightarrow p \models p$ (LN).

1.1.3 (Contraposition 1) $p \models q \Leftrightarrow Nq \models Np$.

Proof: $Nq \models Np \Leftrightarrow \emptyset \models Np, q$ (LN) $\Leftrightarrow p \models q$ (RN).

1.4 (Material Consistency) $I(X) \subseteq Inc \Rightarrow \sim\exists Y[Y \in I(X)\ \&\ I(Y) \subseteq Inc]$.

That is, if X and Y are materially incompatible, they cannot both be valid in the sense of having only incompatibles that are self-incompatible. This result, we should note, depends on our not being in the degenerate frame in which all sets are incoherent. In that case, all sets are also valid.

Proof: Suppose we are not in the degenerate frame. By definition, $I(X) \subseteq Inc \Leftrightarrow (Z \in I(X) \rightarrow Z \in I(Z))$. So only self-incompatible Y could be incompatible with such an X. But since *everything* is incompatible with a self-incompatible Y, it cannot be that $I(Y) \subseteq Inc$. For instance, X would be a non–self-incompatible counterexample to $I(Y) \subseteq Y$. (If X is self-incompatible, it follows that all sets are incoherent, and hence that we are in the degenerate frame.)

1.5 (Formal Consistency) $(X \notin I(X)$ and $X \models p) \Rightarrow X \not\models Np$.

Proof: By NI and Partition, $X \in I(Np) \Leftrightarrow X \models p$. But then X itself is a counterexample to $X \models Np$, since $X \in I(p)$ and $X \notin I(X)$.

It follows immediately that if $\models p$, then $\not\models Np$ (assuming, again, that we are not in the degenerate frame). So the incompatibility logic of negation is consistent.

Since incoherent sets entail everything, it also follows that $(X \models p$ and $X \models Np) \Rightarrow \forall Y[X \models Y]$.

2 Conjunction

2.1 $Kpq \models p$ and $Kpq \models q$.

Proof: By LK, $Kpq \models Y \Leftrightarrow p, q \models Y$. So $Kpq \models p \Leftrightarrow p, q \models p$, and $Kpq \models q \Leftrightarrow p, q \models q$. But $p, q \models p$ and $p, q \models q$ hold by Weakening.

2.2 $(X \models p$ and $X \models q) \Leftrightarrow X \models Kpq$.

Proof: By RK, $X \models Y$, $Kpq \Leftrightarrow (X \models Y, p$ and $X \models Y, q)$. Letting $Y = \emptyset$, then, $X \models Kpq \Leftrightarrow (X \models p$ and $X \models q)$.

3 Negation and Conjunction Together

3.1 $KpNp \models Y$.

Proof: Immediate from LK and the final observation under 1.5 above.

Anything that satisfies 1.2, 1.3, 2.1, 2.2, and 3.1 and *distributivity* is classical (Boolean) logic.

3.2 (Distributivity): $KpAqr \approx AKpqKpr$.

Proof:

(a) $Axy \approx_{df} NKNxNy$. So $KpAqr \approx AKpqKpr$ iff $KpNKNqNr \approx NKNKpqNKpr$.

(b) First direction: $KpNKNqNr \models NKNKpqNKpr$.

(c) $KpNKNqNr \models NKNKpqNKpr$ iff $KpNKNqNr, KNKpqNKpr \models \emptyset$ (RN).

(d) $KpNKNqNr, KNKpqNKpr \models \emptyset$ iff $KpNKNqNr, NKpq, NKpr \models \emptyset$ (LK) iff $p, NKNqNr, NKpq, NKpr \models \emptyset$ (LK).

(e) $p, NKNqNr, NKpq, NKpr \models \emptyset$ iff $p, NKpq, NKpr \models KNqNr$ (LN) iff $p, NKpr \models Kpq, KNqNr$ (LN).

(f) $p, NKpr \models Kpq, KNqNr$ iff $p, NKpr \models Kpq, Nq$ and $p, NKpr \models Kpq, Nr$ (RK).

(g) $p, NKpr \models Kpq, Nq$ iff $p, NKpr \models Nq, p$ and $NKpr \models Nq, q$ (RK).

(h) $p, NKpr \models Kpq, Nr$ iff $p, NKpr \models Nr, p$ and $p, NKpr \models Nr, q$ (RK).

(i) So, plugging (h) and (g) into (f): $p, NKpr \models Kpq, KNqNr$ iff
  (i) $p, NKpr \models Nq, p$ and
  (ii) $p, NKpr \models Nq, q$ and
  (iii) $p, NKpr \models Nr, p$ and
  (iv) $p, NKpr \models Nr, q$.

(j) Now (i-i), and (i-iii) hold because $p \models p$. (i-ii) holds because $p, \text{NK}pr \models \text{N}q, q$ iff $p, \text{NK}pr, q \models q$ (RN), and $p, \text{NK}pr, q \models q$ because $q \models q$. (i-iv) holds because $p, \text{NK}pr \models \text{N}r, q$ iff $p \models \text{K}pr, \text{N}r, q$ (LN), which, by LK, holds iff $p \models p, \text{N}r, q$ and $p \models r, \text{N}r, q$. $p \models p, \text{N}r, q$ holds because $p \models p$, and $p \models r, \text{N}r, q$ holds since it is equivalent by LN to $p, r \models r, q$ and $r \models r$.

(k) So $p, \text{NK}pr \models \text{K}pq, \text{KN}q\text{N}r$ (f), and by (c), (d), and (e), $\text{K}p\text{NKN}q\text{N}r \models \text{NKNK}pq\text{NK}pr$. QED.

(l) Other direction: $\text{NKNK}pq\text{NK}pr \models \text{K}p\text{NKN}q\text{N}r$.

(m) $\text{NKNK}pq\text{NK}pr \models \text{K}p\text{NKN}q\text{N}r$ iff
   (i) $\text{NKNK}pq\text{NK}pr \models p$ and
   (ii) $\text{NKNK}pq\text{NK}pr \models \text{NKN}q\text{N}r$ (RK).

(n) Unpack (m-i): $\text{NKNK}pq\text{NK}pr \models p$ iff $\models p, \text{KNK}pq\text{NK}pr$ (LN) iff
   (i) $\models p, \text{NK}pq$ and
   (ii) $\models p, \text{NK}pr$.

(o) Unpack (n-i): $\models p, \text{NK}pq$ iff $\text{K}pq \models p$ (RN). So (n-i) holds.

(p) Unpack (n-ii): $\models p, \text{NK}pr$ iff $\text{K}pr \models p$ (RN). So (n-ii) holds.

(q) So (m-i) holds.

(r) Unpack (m-ii): $\text{NKNK}pq\text{NK}pr \models \text{NKN}q\text{N}r$ iff $\models \text{KNK}pq\text{NK}pr, \text{NKN}q\text{N}r$ (LN).

(s) $\models \text{KNK}pq\text{NK}pr, \text{NKN}q\text{N}r$ iff $\text{KN}q\text{N}r \models \text{KNK}pq\text{NK}pr$ (RN).

(t) $\text{KN}q\text{N}r \models \text{KNK}pq\text{NK}pr$ iff
   (i) $\text{KN}q\text{N}r \models \text{NK}pq$ and
   (ii) $\text{KN}q\text{N}r \models \text{NK}pr$.

(u) Unpack (t-i): $\text{KN}q\text{N}r \models \text{NK}pq$ iff $\text{KN}q\text{N}r, \text{K}pq \models \emptyset$ (RN).

(v) $\text{KN}q\text{N}r, \text{K}pq \models \emptyset$ iff $\text{N}q, \text{N}r, \text{K}pq \models \emptyset$ (LK) iff $\text{N}q, \text{K}pq \models r$ (LN) iff $\text{K}pq \models r, q$ (LN). But $\text{K}pq \models r, q$ because $\text{K}pq \models q$, since $q \models q$, by (LK). So (t-i) holds.

(w) Unpack (t-ii): $\text{KN}q\text{N}r \models \text{NK}pr$ iff $\text{KN}q\text{N}r, \text{K}pr \models \emptyset$ (RN) iff $\text{N}q, \text{N}r, \text{K}pr \models \emptyset$ (LK) iff $\text{N}r, \text{K}pr \models q$ (LN) iff $\text{K}pr \models q, r$ (LN). But $\text{K}pr \models q, r$ iff $p, r \models q, r$ (LK), and $p, r \models q, r$ because $r \models r$. So (t-ii) holds.

(x) So, by (t): $\text{KN}q\text{N}r \models \text{KNK}pq\text{NK}pr$, so (m-ii) holds.

(y) By (x) and (q), (m) holds: $\text{NKNK}pq\text{NK}pr \models \text{K}p\text{NKN}q\text{N}r$. QED.

(z) So by (y) and (k): $\text{K}p\text{NKN}q\text{N}r \approx \text{NKNK}pq\text{NK}pr$. So $\text{K}p\text{A}qr \approx \text{AK}pq\text{K}pr$, and distributivity holds. QED.

It follows that N, K behave entirely classically.

Besides A, it is useful to define the conditional $Cpq \approx_{df} NKpNq$. A conditional is valid just in case the corresponding entailment holds:

3.3 $p \models q \Leftrightarrow\, \models Cpq$.

Proof: By the definition of C, $\models Cpq \Leftrightarrow\, \models NKpNq$. By RN, $\models NKpNq$ $\Leftrightarrow KpNq \models \emptyset$. By LK, $KpNq \models \emptyset \Leftrightarrow p, Nq \models \emptyset$. By LN, p, $Nq \models \emptyset \Leftrightarrow p \models q$.

Given 3.3, contraposition across $\models (p \models q \Leftrightarrow Nq \models Np)$, proved in 1.3, has as an immediate consequence contraposition for C:

3.4 (Contraposition 2) $\models Cpq \Leftrightarrow\, \models CNqNp$.

We get detachment (*modus ponens*) as a derived rule:

3.5 (Detachment) $KCpqp \models q$.

Proof: By LK, $KCpqp \models q \Leftrightarrow Cpq, p \models q$. By the definition of C, $Cpq, p \models q \Leftrightarrow NKpNq, p \models q$. By LN, $NKpNq, p \models q \Leftrightarrow p \models KpNq, q$. By RK, $p \models KpNq, q \Leftrightarrow (p \models p, q$ and $p \models Nq, q)$. The first of these hold by Weakening, and the second holds since $p \models Nq, q$ is equivalent to $p, q \models q$ (LN), and $q \models q$.

The following two results may help to impart a better intuitive grasp of the behavior of the connectives. The first vindicates the heuristic reading of "$X \models p, q$" as "X entails p or q."

3.6 $X \models Apq \Leftrightarrow X \models p, q$.

Proof: $X \models Apq \Leftrightarrow X \models NKNpNq$ (definition) $\Leftrightarrow X, KNpNq \models \emptyset$ (RN) $\Leftrightarrow X, Np, Nq \models \emptyset$ (LK) $\Leftrightarrow X \models p, q$ (LN).

The next result generalizes Negation Introduction. Where the latter claims that X is incompatible with $\{Np\}$ just in case X entails p, we now show how X relates to multiple negated formulas. In essence, X is incompatible with $\{Np_1, \ldots, Np_n\}$ just in case X entails (p_1 or ... or p_n).

3.7 $X \cup \{Np_1, \ldots, Np_n\} \in$ Inc iff $X \models p_1, \ldots, p_n$.

Proof: The claim is equivalent to $X, Np_1, \ldots, Np_n \models \emptyset$ iff $X \models p_1, \ldots, p_n$. This latter claim follows by n applications of LN.

4 Modality

The *K* axiom is validated by the incompatibility semantics:

4.1 (K) $\models$ CLC*pq*CL*p*L*q*.

Proof: If we are in the degenerate frame, the result follows. So assume
instead that Ø $\not\models$ Ø.

(a) Since we have already shown in 3.3 that $p \models q \Leftrightarrow \models$ C*pq*, it suffices
    to show LC*pq* $\models$ CL*p*L*q*.
(b) By RL, LC*pq* $\models$ CL*p*L*q* $\Leftrightarrow$ ($\models$ CL*p*L*q* or $\not\models$ C*pq*).
(c) By 3.3, $p \models q \Leftrightarrow \models$ C*pq*, $\models$ CL*p*L*q* $\Leftrightarrow$ L*p* $\models$ L*q*.
(d) By RL, L*p* $\models$ L*q* $\Leftrightarrow$ L*p* $\models$ Ø or $\models$ *q*.
(e) By LL, (L*p* $\models$ Ø or $\models$ *q*) $\Leftrightarrow$ (Ø $\models$ Ø or $\not\models$ *p* or $\models$ *q*) $\Leftrightarrow$ ($\not\models$ *p* or
    $\models$ *q*).
(f) Since $p \models q \Leftrightarrow \models$ C*pq*, $\not\models$ C*pq* $\Leftrightarrow$ $p \not\models q$.
(g) So LC*pq* $\models$ CL*p*L*q* $\Leftrightarrow$ ($\not\models$ *p* or $\models$ *q* or $p \not\models q$).
(h) Suppose not. Then ($\models$ *p* and $\not\models$ *q* and $p \models q$). But we showed in
    1.1.2 of Appendix 1 that if $p \models q$ and $\models$ *p*, then $\models$ *q*. So this is a
    contradiction. So LC*pq* $\models$ CL*p*L*q*.

Since PC with *modus ponens* (and substitution) is validated, and we showed
in 3.4 of Appendix 1 that the rule of necessitation $\models$ *p* $\Rightarrow$ $\models$ L*p* holds, so is
the minimal modal system K.

From this it is easy to show that the T axiom—and hence the modal
system T—is validated:

4.2 (T) $\models$ CL*pp*.

Proof: We also showed in 3.5 of Appendix 1 that L*p* $\models$ *p*, and in 3.3
that $p \models q \Leftrightarrow \models$ C*pq*.

4.3 (S4) $\models$ CL*p*LL*p*.

Proof:

(a) By 3.3, $\models$ CL*p*LL*p* $\Leftrightarrow$ L*p* $\models$ LL*p*.
(b) By RL, L*p* $\models$ LL*p* $\Leftrightarrow$ L*p* $\models$ Ø or $\models$ L*p*.
(c) By LL, L*p* $\models$ Ø $\Leftrightarrow$ Ø $\models$ Ø or $\not\models$ *p*.
(d) By RL, Ø $\models$ L*p* $\Leftrightarrow$ Ø $\models$ Ø or $\models$ *p*.

(e) So, plugging (b) and (c) into (a): $Lp \models LLp \Leftrightarrow (\emptyset \models \emptyset$ or $\not\models p$ or $\emptyset \models \emptyset$ or $\models p)$.

(f) So $Lp \models LLp \Leftrightarrow \emptyset \models \emptyset$ or $(\not\models p$ or $\models p)$. But this second disjunct always holds.

Since the system S4 is just T plus the S4 axiom, the incompatibility semantics validates S4.

4.4 (S5) $\models CMpLMp$.

Proof:

(a) By 3.3, $\models CMpLMp \Leftrightarrow Mp \models LMp$.

(b) $Mp \models LMp \Leftrightarrow NLNp \models LNLNp$, since $Mp \approx NLNp$.

(c) By LN, $NLNp \models LNLNp$ iff $\models LNLNp, LNp$.

(d) By RL, $\models LNLNp, LNp \Leftrightarrow \models LNLNp$ or $\models Np$.

(e) By RN, $\models Np \Leftrightarrow p \models \emptyset$.

(f) So, plugging (d) into (c) and (c) into (b): $NLNp \models LNLNp \Leftrightarrow \models LNLNp$ or $p \models \emptyset$.

(g) By RL, $\emptyset \models LNLNp \Leftrightarrow \emptyset \models \emptyset$ or $\models NLNp$.

(h) By RN, $\models NLNp \Leftrightarrow LNp \models \emptyset$.

(i) By LL, $LNp \models \emptyset \Leftrightarrow \emptyset \models \emptyset$ or $\not\models Np$.

(j) By RN, $\models Np \Leftrightarrow p \models \emptyset$, so $\not\models Np \Leftrightarrow p \not\models \emptyset$.

(k) Plugging (i) into (h) and (h) into (g): $\models NLNp \Leftrightarrow \emptyset \models \emptyset$ or $p \not\models \emptyset$.

(l) Plugging (g) into (f) and (f) into (e): $NLNp \models LNLNp \Leftrightarrow \emptyset \models \emptyset$ or $\emptyset \models \emptyset$ or $p \not\models \emptyset$ or $p \models \emptyset$.

(m) But $p \not\models \emptyset$ or $p \models \emptyset$ always holds, so $NLNp \models LNLNp$ always holds.

So the incompatibility semantics validates S5.

Appendix 3: Basic semantic results in the metatheory of incompatibility logic[13]

The proof of completeness follows the familiar Henkin route. Our primary result is that any S5-consistent set Z is satisfiable, and we show this by extending Z to a maximal consistent set Z^* and reading a model off of Z^*. But there are some twists because of the context in which

---

[13] These proofs are due to Alp Aker.

we are working. The relevant notion of satisfaction is not truth in a model, but rather coherence on a frame. So we do not use Z^* to tell us which sentences are true, but instead use it to tell us which sets of sentences are incoherent; that is, we are looking to find out which sets of sentences are to be considered materially inconsistent. Further, our notion of incoherence is a modal notion, so we are interested in which sets of sentences are *necessarily* materially inconsistent. Thus instead of taking each member of Z^* to assert a truth in the model we construct, we consider only members of Z^* of the form $LNKx_1Kx_2 \ldots Kx_{n-1}x_n$, and we take such sentences to specify the incoherent sets of the desired frame.

1   Consistent Sets Are Satisfiable

1.1   Let Z be a set of S5-consistent sentences in the language L (a set such that it is not the case that $Z \vdash_{S5} KpNp$). Then there is a frame Inc_Z such that Z is coherent on Inc_Z (that is, $Z \notin Inc_Z$).

Proof: Let Z^* be a maximal consistent superset of Z and hence deductively closed. By the familiar proof of the Lindenbaum Lemma, on the assumption that Z is consistent, it is easily verified that Z^* exists and that, for each p, either p or Np is in Z^*.

We introduce a new notational convention for the sake of clarity. Let KX be the conjunction of all x_i in X, that is, KX abbreviates $Kx_1Kx_2 \ldots Kx_{n-1}x_n$. (If $X = \{p\}$ let $KX = p$.)

We define our desired frame Inc_Z as follows: $X \in Inc_Z$ iff there is some finite subset $X' \subseteq X$ such that $LNKX' \in Z^*$.

Note that because conjunction is associative and commutative we can neglect the precise ordering and nesting of the conjuncts in KX' when considering $LNKX' \in Z^*$. In the sequel we use this fact without notice.

1.2 Inc_Z satisfies the frame axioms.

Proof: (Persistence) Immediate from construction.

(KI) $X \cup \{Kpq\} \in Inc_Z$ iff $X \cup \{p, q\} \in Inc_Z$.

Suppose $X \cup \{Kpq\} \in Inc_Z$. If that is so then there is a $X' \subseteq X$ such that for some ordering on X we have $LNKKX'Kpq \in Z^*$.

It follows that $X \cup \{p, q\} \in \text{Inc}_Z$. The converse is proved in the same fashion.

(NI) $X, Np \in \text{Inc}_Z$ iff $X \models p$.

Suppose $X \cup \{Np\} \in \text{Inc}_Z$. We want $X \models p$. So assume $W \cup \{p\} \in \text{Inc}_Z$. We want $W \cup X \in \text{Inc}_Z$. Since $W \cup \{p\} \in \text{Inc}_Z$ there is some finite subset $W' \subseteq W$ such that $\text{LNKK}W'p \in Z^*$. Since $X \cup \{Np\} \in \text{Inc}_Z$ we know there is some finite $X' \subseteq X$ such that $\text{LNKK}X'Np \in Z^*$. Now, $\text{NKK}W'p$ and $\text{NKK}X'Np$ truth-functionally imply $\text{NKK}X'\text{K}W'$, so $\text{LNKK}X'\text{K}W' \in Z^*$, which means $W \cup X \in \text{Inc}_Z$.

Now suppose $X \models p$. We want $X \cup \{Np\} \in \text{Inc}_Z$. Now, we know $\text{LNK}p\text{N}p$ is an S5-theorem and so is in Z^*. Thus $\{p, Np\} \in \text{Inc}_Z$. Since $X \models p$, we then know $X \cup \{Np\} \in \text{Inc}_Z$.

(LI) $X \cup \{Lp\} \in \text{Inc}_Z$ iff $X \in \text{Inc}_Z$ or $\exists Y(X \cup Y \notin \text{Inc}_Z \,\&\, Y \not\models p)$.

Suppose $X \cup \{Lp\} \in \text{Inc}_Z$. We want $X \in \text{Inc}_Z$ or $\exists Y(X \cup Y \notin \text{Inc}_Z \,\&\, Y \not\models p)$. We show the equivalent claim that if $\forall Y(X \cup Y \in \text{Inc}_Z$ or $Y \models p)$, then $X \in \text{Inc}_Z$.

Assume the antecedent and instantiate $\{NLp\}$ for Y. Thus $X \cup \{NLp\} \in \text{Inc}_Z$ or $NLp \models p$. We want to show $X \in \text{Inc}_Z$ for each disjunct.

By the first disjunct and our supposition, we have, unpacking the definitions: there is some finite $X' \subseteq X$ such that $\text{LNKK}X'Lp \in Z^*$ and $\text{LNKK}X'NLp \in Z^*$. Since $\text{NKK}X'Lp$ and $\text{NKK}X'NLp$ truth-functionally imply $\text{NK}X'$, it follows that $\text{LNK}X'$ and so $X \in \text{Inc}_Z$.

Applying NI to our second disjunct gives us $\{NLp, Np\} \in \text{Inc}_Z$. Then $\text{LNKNL}p\text{N}p \in Z^*$, which is equivalent to $\text{LCNL}pp \in Z^*$. This, together with $\text{LCL}pp$, gives us Lp (since $\text{CL}pp$ and $\text{CNL}pp$ truth-functionally imply p), which by S4 gives us LLp. Unpacking our supposition $X \cup \{Lp\} \in \text{Inc}_Z$ gives us $\text{LNKK}X'Lp \in Z^*$, which, with LLp, implies $\text{LNK}X' \in Z^*$ (since Lp and $\text{NKK}X'Lp$ truth-functionally imply $\text{NK}X'$). Thus $X \in \text{Inc}_Z$.

Suppose $X \in \text{Inc}_Z$ or $\exists Y(X \cup Y \notin \text{Inc}_Z \,\&\, Y \not\models p)$. We want $X \cup \{Lp\} \in \text{Inc}_Z$ for each disjunct. If $X \in \text{Inc}_Z$, it follows

immediately that $X \cup \{Lp\} \in \mathrm{Inc}_Z$. If $\exists Y(X \cup Y \notin \mathrm{Inc}_Z \ \& \ Y \not\models p)$, then $\not\models p$ and by NI $Np \notin \mathrm{Inc}_Z$. Then $Lp \notin Z^*$. By maximality of Z^*, $NLp \in Z^*$, which implies $LNLp \in Z^*$ by the S5 axioms. Then $Lp \in \mathrm{Inc}_Z$ and our result follows immediately. Thus Inc_Z as we have defined it is a frame. QED (1.2).

1.3 Our original Z is coherent in Inc_Z.

Proof: If Z is incoherent then there is some finite $Z' \subseteq Z$ such that $LNKZ' \in Z^*$ and so $NKZ' \in Z^*$. But $z'_i \in Z^*$ for each $z'_j \in Z'$, and so $KZ' \in Z^*$, which is a contradiction. QED (1.3).

With 1.3 the proof of 1.1 is complete.

2 Completeness of S5 with respect to the incompatibility semantics

2.1. Suppose $\models_{\mathrm{Inc}} p$ on every frame Inc. Then $\vdash_{\mathrm{S5}} p$.

Proof: If $\models_{\mathrm{Inc}} p$ on every Inc, then $Np \models_{\mathrm{Inc}} \emptyset$ on every Inc, and so by 1.1 it follows that Np is S5-inconsistent, in which case $\vdash_{\mathrm{S5}} p$.

This generalizes in the familiar way:

2.2. Suppose $X \models_{\mathrm{Inc}} p$ on every frame Inc. Then $X \vdash_{\mathrm{S5}} p$.

Proof: If $X \models_{\mathrm{Inc}} p$ on every Inc, then $X \cup \{Np\} \in \mathrm{Inc}$ for every Inc, and so by 1.1 we know $X \cup \{Np\}$ is S5-inconsistent, and so $X \vdash_{\mathrm{S5}} p$.

3 Soundness

We adopt as a convenient formulation of the proof theory of S5 the following sequent calculus.

Axioms:

$$(K) \ \vdash CLCpqCLpLq$$
$$(T) \ \vdash CLpp$$
$$(S5) \ \vdash CNLNpLNLNp$$

Rules:

$$(\text{Identity}) \ p \vdash p$$
$$(\text{Contraction}) \ V, X, X, Y \vdash Z \Rightarrow V, X, Y \vdash Z$$
$$(\text{Weakening}) \ X, V, Y \vdash Z \Rightarrow X, W, V, Y \vdash Z$$
$$(\text{Permutation}) \ X, V, W, Y \vdash Z \Rightarrow X, W, V, Y \vdash Z$$
$$(\vdash N) \ X, p \vdash Z \Leftrightarrow X \vdash Z, Np$$

$(N \vdash)$ $X \vdash p, Z \Leftrightarrow X, Np \vdash Z$
$(\vdash K)$ $X \vdash p, Z$ and $X \vdash q, Z \Leftrightarrow X \vdash Kpq, Z$
$(K \vdash)$ $X, p, q \vdash Z \Leftrightarrow X, Kpq \vdash Z$
(Cut) $X \vdash p, Y$ and $V, p \vdash W \Rightarrow X, V \vdash Y, W$

(Necessitation) $\vdash p \Rightarrow \vdash Lp$

3.1 If $X \vdash_{S5} p$, then $X \models p$ on all frames.

Proof: By induction on proof complexity. That is, we verify that each of the axioms is validated by the incompatibility semantics and that each rule preserves validity. The verifications of the rules can be found in Appendix 1 (see 2.1, 4.1.2, 4.1.1, 4.1.4, 4.1.3, 2.2, and 3.4). The verifications of the axioms can be found in Appendix 2 (see 4.1, 4.2, and 4.4).

3.2 If X is S5-inconsistent, then X is incoherent on all frames.

Proof: If X is S5-inconsistent, then there is some finite $X' \subseteq X$ such that $X' \vdash_{S5} p$ and $X' \vdash_{S5} Np$. But then $X' \models p$ and $X' \models Np$ on all frames. But then $X' \cup \{Np\} \in Inc$ and $X' \models Np$ on all frames, so $X' \in Inc$ for all Inc. Thus $X \in Inc$ for all Inc.

4 Compactness

4.1 If for every finite $Z' \subseteq Z$ there is a frame $Inc_{Z'}$ on which Z' is coherent, then there is a frame Inc_Z on which Z is coherent.

Proof: Suppose not. If Z is not coherent on any frame then by 1.1 it is S5-inconsistent. Then it has some S5-inconsistent subset Z'. By 3.2 Z' is incoherent on all frames, which is a contradiction.

Appendix 4: Representation of consequence relations by incompatibility relations

1 Imputing incompatibility relations from consequence relations[14]

1.1 Preliminary remarks   We assume that we have a consequence relation $\vdash$ whose consequent position is either empty or filled by a single sentence.

---

[14] The original representation theorem was proved by the author. But it has been substantially sharpened, and the proof improved, by Alp Aker. Besides the Defeasibility condition required for completeness, the first proof appealed to four conditions that were sufficient for soundness:

That is, $\vdash$ is a relation between sets of sentences and single sentences (or the empty set) for some language L. (We consider other types of consequence relations below; see section 6.) This may be a *material* consequence relation, if the sentences do not have any internal logical complexity (or if we are ignoring what they do have), or it may be a *logical* consequence relation, perhaps defined axiomatically, or by a natural deduction system, or by a sequent calculus.

The Representation Theorem for turnstile $\vdash$ has two conditions:

- General Transitivity: $\forall X, Y \subseteq L \, \forall p, q \in L[(X \vdash p \,\&\, \{p\} \cup Y \vdash q)$
$\rightarrow X \cup Y \vdash q]$.
- Defeasibility: $\forall X \subseteq L \, \forall p \in L[\sim(X \vdash p) \rightarrow \exists Y \subseteq L[\forall q \in L[\{p\}$
$\cup Y \vdash q] \,\&\, \exists q \in L[\sim(X \cup Y \vdash q)]]]$.

Note that General Transitivity has Pure Transitivity as a special case, where Pure Transitivity is:

- Pure Transitivity: $\forall X \subseteq L \forall p, q \in L[(X \vdash p \,\&\, \{p\} \vdash q) \rightarrow X \vdash q]$.

We simply take $Y = \emptyset$.

1.2 Representation definitions

(i) $\mathrm{Inc}(X)$ iff $\forall p \in L[X \vdash p]$.

The basic idea is to read off an incoherence relation from the consequence relation by taking the incoherent sets to be the ones that have *everything* as their consequence. If we start with a *logical* consequence relation, generated by a logic that has *ex falso quodlibet* as a basic or derived rule, this will just be the inconsistent sets: the ones that have as a consequence some sentence and its negation.

We then define incompatibility from incoherence in the usual way:

(ii) $X \in \mathrm{I}(Y)$ iff $\mathrm{Inc}(X \cup Y)$.

(i) Reflexivity: $\forall X \subseteq L[X \vdash X]$.
(ii) Transitivity: $\forall X, Y, Z \subseteq L[(X \vdash Y \,\&\, Y \vdash Z) \rightarrow X \vdash Z]$.
(iii) Monotonicity: $\forall X, Y, Z \subseteq L[(X \vdash Y \,\&\, X \subseteq Z) \rightarrow Z \vdash Y]$.
(iv) Amalgamation: $\forall X, Y, Z \subseteq L[(X \vdash Y \,\&\, X \vdash Z) \rightarrow X \vdash Y \cup Z]$.

Aker showed that, although these conditions are indeed sufficient for the imputed incompatibility relation to generate a semantic consequence relation $\models$ that would hold whenever the original consequence relation $\vdash$ did, they were not in fact necessary for that result. He showed further that General Transitivity is both necessary and sufficient. The "Converse Results" presented below are also due to Aker.

And also define the incompatibility-consequence relation as usual:

(iii) $X \models_I p$ iff $\forall Z \subseteq L[Z \in I(p) \to Z \in I(X)]$.

1.3 Soundness and completeness

> **Representation Theorem** $\vdash$ is sound and complete with respect to $\models_I$ if, and only if, $\vdash$ satisfies General Transitivity and Defeasibility.

We first give the proof from right to left. That is, we show soundness and completeness assuming General Transitivity and Defeasibility.

3.1 (Soundness) If $X \vdash p$, then $X \models p$.

Proof: Suppose not. Then $X \vdash p$ but not $X \models p$ for some X and p. Then by definition of $\models$ there is some $Z \in I(p)$ while $Z \notin I(X)$. Unpacking definitions we have $\forall q \in L[\{p\} \cup Z \vdash q]$ and $\exists r \in L[\sim(X \cup Z \vdash r)]$. Choose some such witnessing r so that $\sim(X \cup Z \vdash r)$. Instantiating $\forall q \in L[\{p\} \cup Z \vdash q]$ we also know $\{p\} \cup Z \vdash r$. Since $X \vdash p$, it follows from General Transitivity that $X \cup Z \vdash r$, which is a contradiction.

3.2 (Completeness) If $X \models p$, then $X \vdash p$.

Proof: Suppose not. Since $\sim(X \vdash p)$ we know by Defeasibility that there are V and r such that $\forall q \in L[\{p\} \cup V \vdash q]$ and $\sim(X \cup V \vdash r)$. Since $X \models p$ we have $\forall Z \subseteq L[Z \in I(p) \to Z \in I(X)]$. Unpacking the definition further we have $\forall Z \subseteq L[\forall q \in L(\{p\} \cup Z \vdash q) \to \forall q \in L(X \cup Z \vdash)]$. Instantiating with $Z = V$ we have $\forall q \in L(\{p\} \cup V \vdash q) \to \forall q \in L(X \cup V \vdash q)$. By modus ponens we have $\forall q \in L(X \cup V \vdash q)$. Instantiating with $q = r$ we have $X \cup V \vdash r$, which contradicts $\sim(X \cup V \vdash r)$.

1.4 Converse results We now show that $\vdash$ satisfies General Transitivity and Defeasibility, assuming $\vdash$ is sound and complete with respect to $\models$.

4.1 (General Transitivity) $\forall X, Y \subseteq L \forall p, q \in L[(X \vdash p \ \& \ \{p\} \cup Y \vdash q) \to X \cup Y \vdash q]$.

Proof: Suppose $X \vdash p$ and $\{p\} \cup Y \vdash q$. By soundness $X \models p$ and $\{p\} \cup Y \models q$. We show $X \cup Y \models q$. To show this, we need to show

that $V \in I(q)$ implies $V \in I(X \cup Y)$. So suppose $V \in I(q)$. Since $\{p\} \cup Y \models q$, this implies that $V \in I(\{p\} \cup Y)$, which implies $Inc(V \cup \{p\} \cup Y)$, which in turn implies $V \cup Y \in I(p)$. Since $X \models p$, we then know that $V \cup Y \in I(X)$. This is equivalent to $Inc(V \cup X \cup Y)$, which is in turn equivalent to $V \in I(X \cup Y)$. Hence $X \cup Y \models q$. By completeness, $X \cup Y \vdash q$.

4.2 (Defeasibility) $\forall X \subseteq L \; \forall p \in L[\sim(X \vdash p) \rightarrow \exists Y \subseteq L[\forall q \in L[\{p\} \cup Y \vdash q] \; \& \; \exists q \in L[\sim(X \cup Y \vdash q)]]]$.

Proof: Suppose $\sim(X \vdash p)$. By completeness, $\sim(X \models p)$. Unpacking the definition of $\models$, we have $\exists Y[Y \in I(p) \; \& \sim(Y \in I(X))]$. Unpacking the definitions of $Y \in I(p)$ and $\sim(Y \in I(X))$, we have $\forall q \in L[\{p\} \cup Y \vdash q] \; \& \; \exists q \in L[\sim(X \cup Y \vdash q)]$.

1.5 Discussion   We have shown that General Transitivity and Defeasibility are jointly equivalent to soundness and completeness. As noted, this initially looks like an ideal result. But the reader might have noticed that our proofs allow for a more precise characterization of the logical relation between these four properties. Put briefly, the situation is this:

   Completeness if, and only if, Defeasibility.

That is, we appealed only to Defeasibility in the proof of completeness, and vice versa. One might expect, then, that we would have:

   Soundness if, and only if, General Transitivity.

But a look at the proofs reveals instead that we have:

   General Transitivity implies soundness.
   Soundness and completeness imply General Transitivity.

We have not been able to eliminate an appeal to completeness in the proof of General Transitivity.

1.6 Generalizations   Having identified the incoherent sets and the semantic entailments, we could proceed to reason logically in the language L. The rules of incompatibility logic are not directly applicable because those rules in general require that the consequents of $\models$ can be sets of formulae, not just single sentences. But having identified the incoherent sets, the entailment

relation we have been using in previous sections is perfectly well-defined. Of course, when we allow claims such as $X \models Y$ we will not have a corresponding consequence $X \vdash Y$ because of expressive limits on $\vdash$.

In cases where $\vdash$ is more expressive, in the sense of allowing multiple formulae in consequent position, there is still a representation result, but the conditions must be slightly different. If $X \vdash \{\gamma_1, \ldots, \gamma_n\}$ has the meaning "X implies γ_1 and ... and γ_1" then the conditions for representation are:

- General Transitivity: $\forall X, Y, W, Z \subseteq L[(X \vdash Y \ \& \ Y \cup W \vdash Z) \rightarrow X \cup W \vdash Z]$.
- Defeasibility: $\forall X, Y \subseteq L[\sim(X \vdash Y) \rightarrow \exists Z \subseteq L[\forall W \subseteq L[Y \cup Z \vdash W] \ \& \ \exists W \subseteq L[\sim(X \cup Z \vdash W)]]]$.

And we must, naturally, also adjust our representation definitions:

(i) $\mathrm{Inc}(X)$ iff $\forall U \subseteq L[X \vdash U]$.
(ii) $X \in \mathrm{I}(Y)$ iff $\mathrm{Inc}(X \cup Y)$.
(iii) $X \models_\mathrm{I} Y$ iff $\forall Z \subseteq L[Z \in \mathrm{I}(Y) \rightarrow Z \in \mathrm{I}(X)]$.

With these changes the representation theorem again holds. Indeed, the proofs require only trivial modification.

If $\vdash$ is instead a disjunctive-consequent turnstile (that is, $X \vdash \{\gamma_1, \ldots, \gamma_n\}$ has the meaning "X implies γ_1 or ... or γ_1"), then the conditions are again different. We can retain Defeasibility as in the previous case, but our transitivity condition becomes:

- General Transitivity: $\forall X, W, Z \subseteq L \ \forall \gamma_1, \ldots, \gamma_n \in L[(X \vdash \gamma_1, \ldots, \gamma_n \ \& \ \{\gamma_1\} \cup W \vdash Z \ \& \ ... \ \& \ \{\gamma_n\} \cup W \vdash Z) \rightarrow X \cup W \vdash Z]$.

The representation definitions are also as in the previous case, except for the definition of entailment:

(iii) $X \models_\mathrm{I} Y$ iff $\forall Z \subseteq L[Z \in \bigcap_{p \in Y} \mathrm{I}(p) \rightarrow Z \in \mathrm{I}(X)]$.

The proofs again require only obvious changes.

2 Discussion of some logical consequence relations

All the logics we consider satisfy General Transitivity. For, as pointed out in the text, that condition is equivalent to the Cut rule:

$$\frac{\Gamma : A \text{ and } \Delta, A : B}{\Gamma, \Delta : B}.$$

This will hold as a derived rule in any system that permits the argument:

$$\frac{\Delta, A : B}{\Delta : A \rightarrow B}$$

by '→' Intro, and then:

$$\frac{\Gamma : A \text{ and } \Delta : A \rightarrow B}{\Gamma, \Delta : B}$$

by '→' Elimination.

And all except relevance logics include *ex falso quodlibet*, equivalent to the rule:

$$\frac{\Gamma : A \text{ and } \Gamma : \sim A}{\Gamma : B}$$

which is required to identify incoherent sets on the basis of their role in the consequence relation, so as to impute the incompatibility relation which in turn determines the incompatibility-entailments.

So the significant condition to consider with respect to various logical consequence relations is Defeasibility.

2.1 Classical logic   It is easy to show that classical logic *does* satisfy Defeasibility. For in this context, defeasibility just comes to the condition that if it is not the case that $p \vdash q$, then there is something that is *inconsistent* with with q and not with p. If q is not a consequence of p, there must be some interpretation on which p is true and q is not true. Pick one such. Now it might, or it might not, be the case that p and q are incompatible or inconsistent (that is, that $\forall U[\{p\} \cup \{q\} \vdash U]$). If they are *not* incompatible, then $\sim q$ is incompatible with q (that is $\forall U[\{q\} \cup \{\sim q\} \vdash U]$) and not with p. If p and q *are* incompatible, then p itself is something that is incompatible with q and not with p—unless p were itself incoherent (= inconsistent: $\forall U[p \vdash U]$), in which case $p \vdash q$, contrary to our hypothesis. So if q is not a logical consequence of p, then there is something that is incompatible with q and not with p, which is the defeasibility condition.

2.2 Modal logics   Most familiar modal logics, including T (sometimes called 'M'), K, B, S4 and S5 (indeed, all the Lewis systems), and many less familiar ones (such as Boolos's GL modal logic of provability) contain all the theorems of the classical propositional calculus PC.

Defeasibility and the arguments and constructions concerning it depend only on the effects of classical negation on the logical consequence relation, so they go through straightforwardly for all normal modal logics.

2.3 Intuitionism   The consequence relation of intuitionistic logic does not satisfy defeasibility. It is the case that whenever an intuitionistic consequence is a good one, everything incompatible (here, inconsistent) with the consequent is incompatible with the antecedent. (That much follows from the soundness result, which depends only on Cut.) But it is not the case that wherever an intuitionistic consequence *fails* there is something that is inconsistent with the consequent but not the antecedent. For instance, it is characteristic of intuitionism that although $\neg\neg p$ does follow from p, p is not a consequence of $\neg\neg p$. The Defeasibility condition requires that there be a 'witness' of the badness of this inference, in the form of something incompatible with p, but not with $\neg\neg p$. In this setting, what is incompatible with p is what is inconsistent with it, and that is whatever entails $\neg p$. But everything that entails $\neg p$ is inconsistent *both* with p *and* with $\neg\neg p$. So there can be no such witness. So Defeasibility fails for the consequence relation of intuitionistic logic. Indeed, the cases where it fails, the non-consequences that fail to have the witnesses incompatibility-defeasibility demands, are just those classical inferences that do not hold good in the intuitionist setting. So intuitionism can be characterized precisely by the cases in which incompatibility-defeasibility fails.

Since the second condition of the representation theorem proved above does not hold for intuitionism, the intuitionistic logical consequence relation is not fully captured by the incompatibility-consequence relation implicit in it. Does that mean that the intuitionistic propositional calculus (and its modal extensions such as intuitionistic S4) does *not* have PC + S5 as its consequence-intrinsic logic? That conclusion would be hasty. For the intuitionistic notion of negation defines a notion of inconsistency that when made to play the role of incompatibility generates a standard incompatibility-consequence relation: that is, one whose proper elaborated-explicating (LX) implicit logical vocabulary is PC + S5. It follows that the techniques introduced here show that alongside the logical consequence relation explicitly defined by the axioms, natural deduction rules, or sequents of intuitionistic propositional calculus, there is *another* logical consequence relation implicitly put in play by the relation of intuitionistic

inconsistency defined by intuitionistic negation. Defeasibility *does* hold for that one, and it permits the introduction of the classical connectives plus the $S5$ modal connectives, by the means outlined in these appendices. In this somewhat extended sense, $PC + S5$ *is* the intrinsic logic of intuitionism (and its modal extensions such as intuitionistic $S4$) too.

To these considerations we may add another, which may be instructive in comparative assessments of the expressive power of intuitionistic versus classical logical connectives (the issue that supersedes concern over which is the *true* or *correct* logic, on the expressive view of the demarcation of logical vocabulary pursued here). If we look at small finite numbers of propositions—say n atomic propositions, along with some, but not all of their negations, and some, but not all of the conditionals relating them—it will often happen that for some incompatibility interpretations (even those that respect the meanings of the connectives to the extent possible, for instance by ensuring that any set containing $\sim p$ is incompatible with any containing p), some inferences we take to be bad ones are endorsed, because everything incompatible with the consequent is incompatible with the antecedent. Intuitively, this is because there just are not *enough* propositions—not enough, that is, to provide witnesses, incompatibility-defeasors, for all the bad inferences. Throwing in some more propositions, for instance, adding more negations, more conditionals, negations of conditionals, and so on, provides the desired defeasors. As n gets larger, and as we more completely form the logical compounds of those atomic propositions, the incompatibility-consequence relation converges on the intuitively—and logically—correct one. One might think of the situation with the two consequence relations generated by intuitionistic logic—the one it defines directly and the one generated by its notion of inconsistency—along these lines. The intuitionistic consequence relation tells us that some consequences are bad, that they do not hold: paradigmatically, the inference from $\neg\neg p$ to p. But while for *most* of the consequences that fail in the intuitionistic setting (for instance, that from $p \vee q$ to p & q) it is possible to give *reasons* justifying the claim that the inference is a bad one, in the form of inconsistency-defeasors, sets of claims that are inconsistent with the conclusion but not with the premises, for *some* (indeed, for just those whose failure distinguishes intuitionistic from classical logic), it is *not* possible to formulate such defeasors, to give reasons of that kind. From the incompatibility point of view—and keeping in mind the way failures to

yield incompatibility-defeasors for intuitively bad inferences can be seen to be due to the expressive impoverishment of systems with "too few" propositions—the failure of Defeasibility for what we may call the 'official' consequence relation of intuitionistic logic amounts to an admission of expressive impoverishment. The intuitionistic logical vocabulary does not have the expressive power to formulate defeasors that could serve as witnesses, as reasons for denying the goodness of inferences the logic nonetheless insists are bad. Such reasons can be given for some of the inferences it rejects (those that are rejected also by classical logic), but not for the rest.

Again, from this point of view, intuitionistic logic shows itself to be incomplete. To defeasor-complete a system containing intuitionistic negation, one would want to add another kind of negation, so contrived that it would supply defeating witnesses inconsistent with the conclusions but not the premises of the inference-forms intuitionism characteristically rejects: paradigmatically, a kind of negation of p (which could be neither intuitionistic nor classical negation) inconsistent with p but not with $\neg\neg p$. Intuitionistic negation provides defeasors only for inferences rejected by *classical* logic. What stands to intuitionistic negation in this respect as it stands to classical negation (which of course is already in equilibrium in the sense of being defeasor-complete)?

Notice that nothing in this discussion of relations between the consequence relations of intuitionistic and classical logic requires the appeal to notions of <u>truth</u>, or <u>truth-value</u>, or <u>bivalence</u>. The difference in the contribution of the two different sorts of negation to the consequence relation is adequately characterized entirely in terms of the notion of logical incompatibility, in the form of inconsistency, that they codify. From the point of view of the pragmatic expressive approach to the demarcation of logical vocabulary pursued here, understanding those negations is a matter of understanding which aspects of *material* incompatibility they make explicit.

2.4 Relevance logic   Relevance logic claims as its primary, in some sense characteristic, virtue its *rejection* of the principle *ex falso quodlibet*. It is just this principle on which the construction offered here of standard incompatibility relations from standard consequence relations turns. For the incoherent sets (and hence the incompatibility relations between sets) are defined as those that count *everything* among their consequences. Now,

there is nothing sacred or inevitable about this procedure. What is needed to get the enterprise off the ground is *some* way of picking out sets of sentences that are incoherent, in terms of their distinctive role in the consequence relation. In the present case, where what is at issue is a *logical* consequence relation, <u>incoherence</u> amounts to <u>inconsistency</u>. So the question that must be addressed to relevance logic is: What property, expressible entirely in terms of the consequence relation, distinguishes *inconsistent* from *consistent* sets of sentences? It will *not* be, of course, that the inconsistent ones are those that have *everything* as their consequences. But if not that property, what does distinguish the inconsistent sets in the context of the consequence relation of relevance logic? The unsettling answer, at least for the pure arrow fragment of R and E—the philosophical core and paradigm of the enterprise—is: nothing. There is no way at all to distinguish inconsistent from consistent sets purely on the basis of their role in the consequence relation. (Of course, one can always pick them out as the sets that contain or entail both *p* and *not-p*. But the point is that one cannot recover *that* information just from how the sets behave as premises or conclusions of logically good inferences.)

Now, even if there *were* some way of picking out the relevantly inconsistent sets, it would still have to be shown, and might not be true, that, when treated as the incoherent sets, they would define a *standard* incompatibility relation. (It's a tough counterfactual.) But one might be given pause by the fact that as far as the consequence relation of relevance logic is concerned, consistent and inconsistent sets are indistinguishable. That seems like a Bad Thing, suggesting that this way of recoiling from *ex falso quodlibet* has somehow gone too far. It is a defect that can be remedied. Various ''impure'' forms of relevance logic introduce special notions of and signs for *absurdity*—none of which, of course, automatically have the consequence of entailing everything—which do permit the discrimination of inconsistent sets as those that have absurdity among their consequences. I've not looked at those to see which, if any, might generate standard incompatibility relations.

This observation about the peculiar inability of relevance logic to discriminate inconsistent sets of sentences solely on the basis of their behavior as premises or conclusions of logically good inferences raises the general question of what one *can* tell about the logical form of sentences in virtue of their role in logical consequence relations, in the context of

different logics. It is easy to show that, under natural assumptions, it is possible to recover the logical form of individual sentences from the role they play as premises and conclusions in the *classical* logical consequence relation, *up to negation*. That is, first, because p is logically equivalent to $\sim\sim p$, one cannot tell sentences of those two forms apart. And further, one cannot in general tell p and $\sim p$ apart—not in the sense that they have the *same* role in the consequence relation (after all, $p \vdash p$ and not $p \vdash \sim p$), but in the sense that systematically substituting p for $\sim p$, or vice versa, makes no difference to the consequence relation. On the other hand, in the *intuitionistic* setting, one can not only tell p from $\sim\sim p$, there is also a systematic asymmetry between the consequences and consequential antecedents of p and $\sim p$. The intuitionistic logical consequence relation is *categorical* for the logical form of the sentences it relates, in the sense that it suffices fully to determine their logical form. From an inferentialist semantic perspective, this feature amounts to a significant expressive advantage of intuitionist over classical (not to mention relevance) logic.

6

Intentionality as a Pragmatically Mediated Semantic Relation

1 Pragmatism and semantics ("I sing of words and the world.")

Under the banner of "analytic pragmatism" I have been illustrating how deploying the metavocabulary of meaning-use analysis can both broaden our understanding of possible kinds of semantic analysis and help turn contemporary pragmatism from a primarily critical into a more constructive instrument—from a weapon suitable for the heavy, heroic, but occasional work of slaying dragons of conceptual confusion into a tool adapted for everyday domestic analytical and theoretical use. As its name implies, the broader sort of analysis I have been recommending considers relations between meaning and use (between vocabularies and practices-or-abilities). The principal complex resultant meaning-use relation I have focused on is that which obtains when one vocabulary is algorithmically elaborated from and explicating of ("LX for" in short) some practices-or-abilities that are PV-sufficient to deploy another vocabulary. This, I have argued, is the genus of which *logical* vocabulary is a species. In Lecture 4 I showed how modal vocabulary and the vocabulary used to discuss specifically conceptual norms can also be understood and introduced as like logical vocabulary in being *universally* LX: elaborated from and explicating of practices PV-necessary for *every* autonomous discursive practice. Furthermore, the technical means employed in meaning-use analysis, for instance, both algorithmic decomposition and complex meaning-use relations, are not only *analogous* to the logical relations appealed to by the semantic logicist dimension of classical philosophical analysis, but are important for understanding why logical vocabulary deserves

the special role it has traditionally been taken to play in such semantic analysis.

It is high time, however, to look more closely at the claim that the topic I have been addressing deserves to be called *semantic* analysis in the first place. I have not so far said anything at all about word–world relations, or about representation. And the formal incompatibility semantics I presented last time is notable in part precisely for the fact that it does *not* deploy a notion of <u>truth</u>. I have, to be sure (as promised in the title of the lectures) talked about saying and doing, and about some of their relations: about the kinds of doings that are sayings, and about the kinds of sayings that specify those kinds of doing. But my talk of 'vocabularies' and the practices of deploying them can make it look as if *all* that is in play is words and their use. If the *world* is left out of the story, what justification could there be for saying that *meaning* has not been? And if a slide has been initiated in lining up saying and doing with meaning and use, it would seem only to be accelerated by my practice of talking about both in terms of the distinction between semantics and pragmatics. The use of the term 'pragmatics' to encompass meaning-conferring aspects of use in general is non-standard, though I think it is fairly straightforwardly motivatable. But in what sense, it might be asked, have the *meanings* of the vocabularies I have addressed been under discussion, if the relations between those *words* and the *world* that using them to *say* something consists in talking *about* do not come into view? Doesn't the story I have been telling remain too resolutely on the 'word' side of the word/world divide?

The short answer is that while at least some kinds of representings (for instance linguistic utterance-tokenings thought of as mere sign-designs, items or events in the natural order, Wittgenstein's signpost considered just as a piece of wood) can intelligibly be specified independently of what they represent, when properly conceived, practices and abilities are *not* the kind of thing that can be separated from the objects they involve in the way necessary for them to fall on the 'word' side of a word/world gulf. Engaging in discursive practices and exercising discursive abilities is using words to say and mean something, hence to talk about items in the world. Those practices, the exercise of those abilities, those uses, *establish* semantic relations between words and the world. This is one of the big ideas that traditional pragmatism brings to philosophical thought about semantics: don't look, to begin with, to the relation between representings

and representeds, but look to the nature of the doing, of the process, that institutes that relation. It is an idea that is explicit in Dewey, and at least implicit in Wittgenstein. This pragmatist privileging of process over relation in the order of semantic explanation is worth looking at more closely.

It can be thought of in terms of a nested sequence of claims:

P1] A founding idea of pragmatism is that the most fundamental kind of intentionality (in the sense of directedness towards objects) is the *practical* involvement with objects exhibited by a sentient creature dealing skillfully with its world.

P2] The most basic form of such activity is a Test-Operate-Test-Exit (TOTE) cycle of perception, performance, assessment of the results of the performance, and further performance—that is, a process or practice consisting of an open-ended sequence of feedback-governed performances.[1]

It includes both what a predator does in stalking its prey and what a builder does in constructing a house.

As we have seen, often a practice-or-ability in this sense can show up as *complex*, in that a suitably rich VP-sufficient metavocabulary can specify it as the algorithmic elaboration of more basic reliable differential responsive dispositions. Doing so displays its structure as comprising a sequence of simpler doings, as exhibiting a plan structure, or as implementing a conditional branched-schedule algorithm. Feedback-governed processes, practices, and abilities exhibiting this sort of complexity cannot in principle be specified without reference to the changes in the world that are both produced by the system's responses and responded to within each loop in the TOTE cycle. This fact underlies another important pragmatist claim:

P3] Feedback-governed practices are 'thick', in the sense of essentially involving objects, events, and worldly states of affairs. Bits of the world are *incorporated* in such practices, in the exercise of such abilities.

In this regard they contrast with words and sentences, considered merely as sign-designs or items in the natural world, which are 'thin' in that they

---

[1] I don't say that this stripped-down description is *sufficient* to pick something out as sentient—only that sentient things must satisfy this abstract characterization.

can be specified independently of a specification of the objects or states of affairs they refer to or represent. This difference is (I think properly) put forward as one of the cardinal advantages of approaching semantics from a pragmatist direction. As I have indicated, I think it, too, should be understood in terms of features of the vocabularies that are VP-sufficient to specify the practices in question. Think of the practices of attaching two flat objects by using nails and a hammer (henceforth ''hammering'') or screws and a screwdriver (henceforth ''screwing''). You cannot say what hammering and screwing are without referring to the actual objects incorporated in them in different ways: the hammers, nails, and so on, that play essential roles in those practices. This is a VV-necessity relation concerning the vocabularies that are VP-sufficient to specify this basic sort of practical transaction: in order to specify this kind of practice-or-ability, one must use vocabulary that picks out *objects* they involve.

The next piece of the pragmatist approach to intentionality is the claim that:

P4] The specifically *semantic* intentionality displayed in language-use, engaging in *discursive* practices, deploying an autonomous *vocabulary*, should be understood both as a development of and as a special case of the sort of basic *practical* intentionality exhibited already by the kind of feedback-governed transactions mentioned in the first three theses.

Now we must ask what the relation is between understanding saying as that sort of doing, on the one hand, and understanding it as representing—as establishing a *semantic* relation between subjective doings and objective states of affairs, between representings and representeds—on the other.

It is only in terms of an answer to this question that we can give a definite sense to the final claim of the pragmatist line of thought I have been sketching:

P5] One cannot understand the two poles of specifically semantic or discursive intentionality—knowing and acting subjects and the objects they know of and act on, their representing activities and the objects and objective states of affairs they represent—independently of the semantic intentional relations they stand in to one another, and then somehow bolt together those ways of understanding the

relata to understand those relations between them. One must rather start with an understanding of the thick, essentially world-involving practices engaged in and abilities exercised, and abstract from or dissect out of that an understanding of the two poles of the semantic intentional relations those practices and abilities institute or establish.

It is commitment to this order of semantic explanation that is, I think, most characteristic of the philosophical tradition I have been calling "pragmatist."[2]

2 Normative and modal vocabularies again

How can the metavocabulary of meaning-use analysis I have been developing for analyzing complex resultant meaning-use relations be applied to make more definite the Deweyan claim about the possibility of extracting an understanding of the relata of intentional and semantic relations from a

---

[2] The theme is pervasive in Dewey's writings, from the time of his early "The Reflex-Arc Concept in Psychology" (1896; reprinted in *John Dewey: The Early Works 1882–1898*, Jo Ann Boydston [ed.], Southern Illinois University Press, 1972, vol. 5, 97–110). Here are some representative passages from the late *Experience and Nature* (1925; reprinted in *John Dewey: The Later Works 1925–1953*, Jo Ann Boydston [ed], Southern Illinois University Press, 1981, vol. 1):

[By contrast to the traditional view:] Subjective and objective distinguished as factors in a regulated effort at modification of the environing world have an intelligible meaning. (p. 185)

We begin by noting that "experience" is what James called a double-barrelled word. Like its congeners, life and history, it includes what men do and suffer, what they strive for, love, believe and endure, and also how men act and are acted upon, the ways in which they do and suffer, desire and enjoy, see, believe, imagine—in short, processes of experiencing. "Experience" denotes the planted field, the sowed seeds, the reaped harvests, the changes of night and day, spring and autumn, wet and dry, heat and cold, that are observed, feared, longed for; it also denotes the one who plants and reaps, who works and rejoices, hopes, fears, plans, invokes magic or chemistry to aid him, who is downcast or triumphant. It is "double-barrelled" in that it recognizes in its primary integrity no division between act and material, subject and object, but contains them both in an unanalyzed totality. "Thing" and "thought," as James says in the same connection, are single-barrelled; they refer to products discriminated by reflection out of primary experience. (p.19)

I still believe that on theoretical, as distinct from historical, grounds there is much to be said in favor of using "experience" to designate the inclusive subject-matter which characteristically "modern" (post-medieval) philosophy breaks up into the dualisms of subject and object, mind and the world, psychological and physical. (p. 362)

The value of experience for the philosopher is that it serves as a constant reminder of something which is neither exclusive and isolated subject or object, matter or mind, nor yet one plus the other. The fact of integration in life is a basic fact, and until its recognition becomes habitual, unconscious and pervasive, we need a word like experience to remind us of it, and to keep before thought the distortions that occur when the integration is ignored or denied. (p. 385)

conception of practices that conditionally link sequences of perception and action in processes of transaction with an environment?

The way I will pursue here looks to our earlier discussion of the expressive roles characteristic of *normative* and *modal* vocabularies. I have made a number of claims about them over the course of these lectures. The most basic of these, introduced in the fourth lecture, was that each of these vocabularies should be seen as LX for, as elaborated from and explicitating of, various features essential to every autonomous discursive practice. The features of discursive practice from which the normative vocabulary of commitment and entitlement is elaborated and which it makes explicit are different from those from which the modal vocabulary of necessity and possibility are elaborated and made explicit. But they are intimately related. What I want to claim now is that those features correspond, respectively, to the subjective and the objective poles of intentional relations. Further, the relation between normative and modal vocabulary explored in my fifth lecture—in particular, the way in which normative vocabulary can be understood to serve as a pragmatic metavocabulary for modal vocabulary—provides an important tool for understanding the relation between the use of expressions as representations and what they represent.

The basic idea is that normative vocabulary makes explicit important features of what knowing and acting subjects *do* when they deploy a vocabulary, when they *use* expressions so as to *say* something. And modal vocabulary makes explicit important correlative features both of what is *said* and of the objective world that is talked *about*. Put another way, normative and modal vocabulary, each in its own way, articulate discursive commitments. But normative vocabulary addresses in the first instance *acts* of committing oneself, whereas modal vocabulary addresses in the first instance the *contents* one thereby commits oneself to—not in the sense of what other *doings* committing oneself to a claim commits one to, but in the sense of how one has committed oneself to the world being, how one has represented it as being.[3] If there is anything to this idea, then thinking about complex, pragmatically mediated resultant semantic

---

[3] We are now going to look at the fine structure that articulates the very broad and capacious relation of VP-sufficiency that had to be put in place to see all of logical, modal, and normative vocabularies as making explicit features of the use of ordinary vocabularies, thus redeeming a promissory note that has been outstanding since my second lecture.

relations between normative and modal vocabularies[4] is a way of thinking analytically both about discursive intentionality (the kind that involves distinctively *semantic* relations), and about the relation between what one who engages in a discursive practice *does* and what she *says about* the objective things she thereby represents or talks about.

In the meaning-use diagram shown in Figure 6.1, the sub-practices of the autonomous discursive practice that are labeled 'subjective-normative' and 'objective-modal' are to be identified as those picked out by the dual conditions that they are the practices-or-abilities from which practices PV-sufficient for the introduction of deontic normative vocabulary (or, respectively, alethic modal vocabulary) can be elaborated, and the practices-or-abilities that are made explicit by that vocabulary in the sense that it is VP-sufficient to specify them. In this way, the complex, resultant meaning-use relations they stand in are used to dissect out what then show up as components of autonomous discursive practices. How might we think about the aspects of discursive practices that are picked out in this way by the dual LX-ness conditions in terms of which the *use* of *normative* and *modal* vocabularies is analyzed?

In the senses in which I have been using the terms, a creature's practical engagement with its world exhibits *practical intentionality* insofar as it is feedback-governed, that is, specifiable (in a sufficiently rich vocabulary) as having an algorithmic TOTE structure in which each cycle is mediated by its differential responses to the effects of its own performances. Specifying the behavior of a system in such terms is taking or treating it as practically *directed toward* the features of its environment that play a suitable dual

---

[4] Put somewhat more carefully, I explore here an intimate sort of connection between (some) deontic modalities and (some) alethic modalities. Only 'some' in the first case, because (for instance) *moral* normativity can also be put in deontic terms, and I am only addressing the *conceptual* variety of normativity: norms governing the application of concepts. And only 'some' in the second case because the alethic modalities (necessities and possibilities) I am discussing are not, or are not restricted to, metaphysical necessities in the Kripkean sense. They include those involved in laws of nature that support counterfactuals that may not be metaphysically, but only physically, necessary. And they include other conceptual necessities such as those involving the incompatibility of color and shape properties that are harder to pin down. (I take it that it is a *geometrical*, rather than a *physical* fact that being rectangular and being circular are incompatible properties of plane figures. And it is not clear how to characterize the incompatibility of *red* and *green*.) The kind of alethic modality (because the kind of modal incompatibility) I am after cuts across a lot of the usual categorizations, because it is in play wherever material inferences have a range of counterfactual robustness. Any such range corresponds to a judgment as to what is and what is not *possible*, in the sense that matters for the kind of *semantic contents* I am concerned to think about vocabulary as expressing.
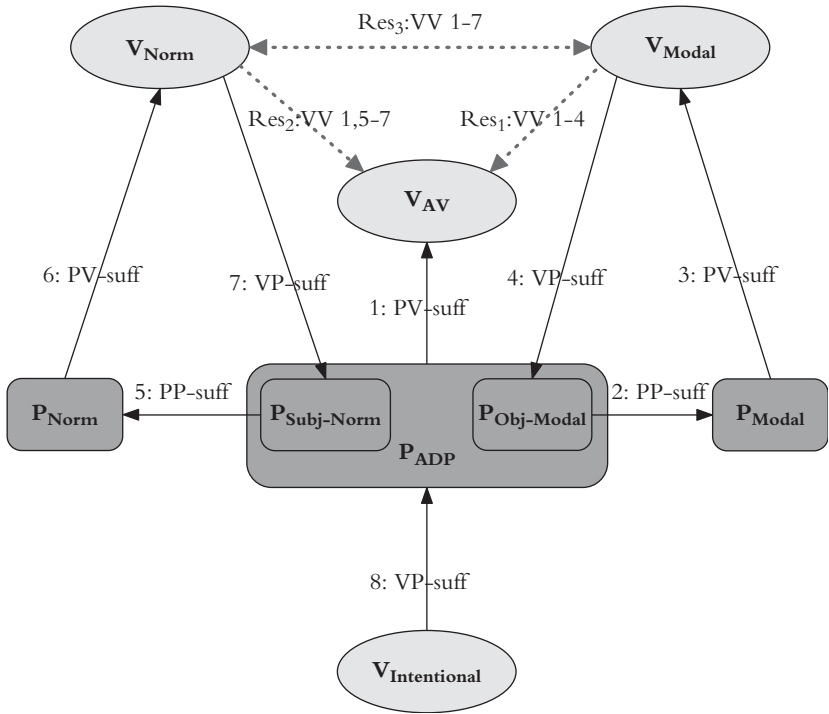
Figure 6.1 Analyzing or dissecting discursive practices into subjective-normative and objective-modal components

role in the reliably covarying causal chains of events that serve as both inputs to and outputs from the system that engages in a process with this structure. Such a system counts as exercising *discursive* abilities, or engaging in discursive practices, hence as exhibiting specifically *discursive* intentionality, insofar as the differential responsiveness of the system to the results of its own performances is essentially mediated by states whose functional role in the feedback process can be understood only by taking them to be *propositionally* contentful, that is, by specifying them in an *intentional* vocabulary—paradigmatically as involving the claim, belief, preference, or intention that *p*, where '*p*' is a declarative sentence in the VP-sufficient intentional vocabulary specifying the practices-or-abilities in question (which may or may not be a sub-vocabulary of the autonomous vocabulary being deployed). I have been conducting this investigation within the scope of the assumption that a necessary element of that requirement is that the process that mediates between differential sensitivity

to the effects of prior performances and differential dispositions to produce subsequent performances—between testing and operating in the TOTE cycle—be governed by and exhibit sensitivity to norms articulating relations of material incompatibility and inferential consequence. Sensitivity to the applicability of such conceptual norms is manifested in the way the system *updates* its beliefs, preferences, and intentions, thereby moving from one functional state to another, during the process of its engagement with its environment. At the beginning of Lecture 5 I pointed out how the deontic normative vocabulary of 'commitment' and 'entitlement' could be used to codify many of the different kinds of material inferential and incompatibility *relations* that structure these inferential *processes, practices*, or *activities*.

The next question, then, is how the sort of directedness at objects via feedback engagement with them that is characteristic of *practical* intentionality, turns into something intelligible as *representation* of those objects when the process of practical engagement takes the form of deontic updating structured by material *inferential* and *incompatibility* relations, that is, when it becomes *discursive* intentionality. Answering that question is beginning to work out the pragmatist's order of semantic explanation. Telling that story requires saying how, within the discursive realm, representational 'of'-intentionality is related to expressive 'that'-intentionality, that is, how what one is talking *of* or *about* (representing) is related to what one *says*, of or about those things. And doing that will enable us to get clearer about the nature of the intimate relation between what it is about our practice of *saying* that is made explicit by *normative* vocabulary and what it is about what is *said* that is made explicit by *modal* vocabulary—which is my suggestion as to how to pursue the pragmatist explanatory aspiration: by describing a complex, resultant meaning-use relation between these vocabularies that offers yet a further way (beyond those considered in Lectures 4 and 5) of filling in and following out Sellars's dark but suggestive remark that "the language of modality is a 'transposed' language of norms."

3 Discursive representation and rational rectification

Consider a non-autonomous vocabulary, a language fragment, centered on the use of the term 'acid'. In the toy practice I am envisaging, if a liquid

tastes sour, one is committed and entitled to apply the term 'acid*' to it. And if one is committed to calling something 'acid*', then one is committed to its turning phenolphthalein blue. I imagine that the community using this term displays wide agreement, under concurrent stimulation, concerning what things are sour and what things are blue, and has experts certifying some vials as containing phenolphthalein. In using the term 'acid*' with these circumstances and consequences of application, the community is implicitly endorsing the propriety of the material inference from a liquid's tasting sour to its turning phenolphthalein blue. If a practitioner comes across a kind of liquid that tastes sour but turns phenolphthalein red, she finds herself with commitments that are materially incompatible, by her own lights. For she infers from its sourness that it is an acid*, and from its being an acid* that the phenolphthalein solution to which it is added is blue. But exercising her reliable differential responsive dispositions directly, she non-inferentially acquires an incompatible observational commitment to the phenolphthalein solution being red. She cannot be entitled to both. Inferential expansion of one observation has led to a commitment incompatible with another. To repair that incompatibility (to update her commitments), she is obliged either to relinquish the claim that the liquid tastes sour, or to relinquish the claim that phenolphthalein solution is red, or to revise her concept of an acid* so that it no longer mediates the inference that caused the problem—perhaps by restricting its applicability to *clear* liquids that taste sour, or by restricting the consequence to turning phenolphthalein blue when the liquid is heated to its boiling point. Entitling oneself to any of these moves involves further commitments it may not be easy to entitle oneself to, and none of them may ultimately be successful. But in any case, something has been learned.

This little parable of one cycle of practical operating and testing involves discursive updating of three basic types: expanding one's observations by drawing commitment- and entitlement-preserving inferential conclusions; registering any resulting materially incompatible commitments; and repairing them by modifying or relinquishing some of those commitments, or the concepts that link them inferentially. Those inferential links between observational concepts, whose applicability can also be elicited non-inferentially by the exercise of reliable differential responsive dispositions, engender the possibility of *friction* between the world and the deployment of vocabulary in a practical cycle of perception-and-performance articulated by those

material inferential and incompatibility relations.[5] In the context of the set of practices-and-abilities I described, the world, by presenting a liquid that tastes sour and turns phenolphthalein red, is telling our imaginary community that it cannot have the concept <u>acid*</u> with the original circumstances and consequences of application. For what has been revealed is that, contrary to the material inference curled up in that concept, it is *not necessary* that sour liquids turn phenolphthalein blue. It is *possible* that a liquid both be sour and turn phenolphthalein red. Where enough discursive updating TOTE cycles of this sort have been engaged in to produce a relatively stable and successful discursive practice, objective facts about what actually follows from and what is incompatible with what will have been incorporated in the material inferences and incompatibilities that articulate the concepts expressed by the vocabulary deployed according to the practical norms implicit in that practice. This essentially holistic process involves getting on to how things objectively are not just by making true claims, but also by acknowledging the right concepts.

We see here in microcosm a pivotal relation between what is expressed by the use of *normative* vocabulary to codify central features of the inferentially articulated doings of knowing and acting discursive *subjects* and what is expressed by the use of *modal* vocabulary to characterize central features of the objective world they talk about and act in: its *laws*, what connections are *necessary*, what is really *possible*. When all goes well, the normative structure of consequential commitments and entitlements, including incompatibilities, tracks the modal structure of laws relating possible facts. Taking an inference to be a good one even in counterfactual circumstances by endorsing an appropriately modally qualified conditional is what one needs to *do* in order to *say that* a law holds objectively.

An essential part of what one is *doing* in committing oneself (doxastically or practically) to some claimable content is taking responsibility for *integrating* it into a whole constellation of such commitments, by following out the inferential consequences it has in the context of its fellows, and subjecting it to rational criticism by confronting it with any concomitant commitments that turn out to be materially incompatible with it. Engaging in that fundamental sort of discursive activity is what Kant called "synthesizing the

---

[5] A useful pragmatist perspective is provided by Huw Price's "Truth as Convenient Friction," *Journal of Philosophy*, 100 (2003), 167–90. Also in Patrick Grim, Gary Mar, and Peter Williams (eds.), *The Philosopher's Annual*, vol. 26 (Ridgeview Publishing, 2003).

transcendental unity of apperception.'' Apperception is discursive (that is, conceptually articulated) awareness: undertaking commitments whose contents can be specified by declarative sentences. The *unity* of those discursive commitments is a *normative* unity: a matter of taking responsibility for one's commitments by acknowledging what else they commit and entitle one to, and what other contents to which one may initially be committed they preclude entitlement to. Discursive updating aims at the material inferential completeness and compatibility of one's commitments, in the normative sense that insofar as one falls short of those ideals, one is normatively obliged to *do* something about it, to repair the failure.

The point I want to focus on, however, is that the account of feedback loops of perception-and-performance normatively governed by relations of material inference and incompatibility is a pragmatic version of Kant's account of synthesizing a *transcendental* unity of apperception. In his terminology, transcendental logic is distinguished from general logic by its concern with the *content*, rather than just the *form*, of judgments. And that content is understood in terms of *representation of objects*. Now, I have urged that an essential element of the propositional contentfulness expressed by declarative sentences and attributed by 'that'-clauses in ascriptions of intentional states using vocabulary such as 'claims that' and 'believes that' consists in those contents standing in material inferential and incompatibility relations to one another. And these are the very relations that normatively govern the discursive updating process I have lined up with Kant's notion of synthesizing a transcendental unity of apperception. But what, we may ask, makes the unity in question deserve to be called *transcendental*, in a sense that invokes representation of objects?

The answer lies in the way in which acknowledging material inferential and incompatibility relations among commitments, essentially involves re-presenting objects as having properties (perhaps complex relational ones) that stand in corresponding relations to one another. In drawing inferences and 'repelling' incompatibilities, one is taking oneself to stand in representational relations to objects that one is talking *about*. A commitment to A's being a dog does *not* entail a commitment to B's being a mammal. But it *does* entail a commitment to A's being a mammal. Drawing the inference from a dog-judgment to a mammal-judgment *is* taking it that the two judgments represent one and the same object. Again, the judgment that A is a dog is not incompatible with the judgment that B is a fox. It *is* incompatible with the

judgment that *A* is a fox. Taking a dog-judgment to be incompatible with a fox-judgment *is* taking them to refer to or represent an object, the *one* object to which incompatible properties are being attributed by the two claims.

Representational purport is accordingly an integral feature of the process of practically acknowledging material inferential and incompatibility relations that is discursive updating. It involves a kind of *triangulation* on the object represented that is evidently the discursive, theoretical culmination of a progression that begins with the way physical objects are incorporated in feedback-governed practical engagements with things. The most basic sort of practical triangulation on objects happens when the *result* of one responsive performance is itself the *stimulus* eliciting a further response—as, for instance, with the nail one hits with the hammer, perceives the effect on, and then responds to anew by hitting it again, until it is driven flush. (And an analogue of this feature of hammering is evidently also exhibited in the case of screwing.) The next level of triangulation includes both the practical exercise of reliable differential responsive dispositions and relations of material inference-and-incompatibility, and depends on their interaction. It might be epitomized by the way the one liquid observed both to taste sour and to turn the phenolphthalein red is caught up in the toy practice of using the term 'acid*' I sketched above. The next level dispenses with the immediate involvement of practical responsiveness entirely, as in the purely inferential-and-incompatibility relations among the concepts <u>dog</u>, <u>mammal</u>, and <u>fox</u>. These forms of triangulation begin with purely practical incorporation of something objective in a feedback-governed process, and end with purely theoretical objective representational purport. **The triangulation that consists in acknowledging material incompatibilities and inferences is, in a nutshell, how the *normative* demand for a *rational* unity of apperception (judgments) makes intelligible *representational* purport: what it is to take or treat judgments in practice as representing or being about objects**. Together, the different stages in this progression illustrate both how *discursive* intentionality is rooted in and grows out of *practical* intentionality, and how, within the discursive realm, *representational* 'of'-intentionality is inseparably related to *expressive* 'that'-intentionality—indeed, is intelligible in terms of it, by means of a pragmatically mediated semantic relation.

Expanding commitments inferentially, noting and repairing incompatibilities—the ampliative and critical dimensions, the inhalation and

exhalation of living discursive activity—*is* synthesizing a normative *rational unity of apperception*, for it is treating one's commitments as *reasons* for and against other commitments. And we have seen how doing that is the endless production and reproduction (at once practical and ideal) of a unity (a holistic rational *system* of commitments) that is *transcendental* in Kant's sense, in that it is the process that institutes *representational* relations. Acknowledging the rational critical responsibility implicit in taking incompatible commitments to oblige one to *do* something, to update one's commitments so as to eliminate the incompatibility, is what one must *do* in order in practice to be taking oneself to be talking *about* or representing things, in the normative sense of making oneself responsible *to* them for the correctness of one's claims and the success of one's actions (the fulfillment of one's intentions, the satisfaction of one's desires).

Discursive intentionality of both sorts, 'that'-intentionality and 'of'-intentionality—indeed, as we can now say, discursive intentionality along both of its inextricably intertwined expressive and representational dimensions—is the paradigmatic *semantic* phenomenon. I have been rehearsing some of the ways in which it can be seen to be a *pragmatically mediated* semantic phenomenon by looking at some of the things one must *do*—the practices one must engage in, the abilities one must exercise—in order to say *of* something *that* it is thus-and-so. Of course, in some sense no one ever doubted the general semantic pragmatist claim: what else but the way it is *used* could make a vocabulary *mean* or *represent* something? But the analytic pragmatist is interested in saying in detail in other (VP-sufficient) terms what one must *do* (what is PV-sufficient) to count thereby as saying or representing something. And the claim that one cannot understand what semantic relations *are* except by understanding the practical processes and activities by which they are instituted (a sort of pragmatic sense-dependence claim) has not always been thought to be a truism (or even just to be true).

4 Two senses of 'incompatible'

The most surprising claim I have made here is that a special insight into the nature of semantic intentional relations and their relation to the discursive practices of rational amplification and rectification (normatively governed respectively by material inferential and incompatibility relations)

that establish and maintain them can be gleaned by looking at complex resultant, pragmatically mediated meaning-use relations between *normative* and *modal* vocabularies. As we have seen, this claim is the final move in a way of working out a pragmatist order of semantic explanation that we can think of as involving four sequential steps. First is the idea that the most basic form of intentionality is feedback-governed practical transactions: TOTE cycles of differential response and response to the effects of the response. To understand a process as having this structure is to take it to be algorithmically decomposable, that is, specifiable as the algorithmic elaboration of more basic differential responsive capacities (a special kind of PP-sufficiency relation). To do that is to specify it in terms of some more basic abilities and the way algorithmic elaborative capacities are deployed to implement the more complex ability (a special kind of VP-sufficiency relation). (This is a very basic kind of *functionalism*—functionalism about practical intentionality—which we have seen to be integral to the pragmatist version of the program of artificial intelligence.) The second move is to understand specifically *discursive* intentionality, the kind that institutes *semantic* relations, as a species of such feedback-governed practical engagement in which performance and response are mediated by relations of material inference and incompatibility. The third stage in this progression is then the claim that, instead of thinking of the intentional nexus to begin with in terms of relations of a distinctive kind ('semantic') between *things* that we can in principle characterize antecedently to and independently of their semantic relation, representeds and representings (as Fodor puts it, horses and 'horse's), we think rather of two dimensions abstracted from or brought into relief within such a feedback-governed process of practical engagement, mediated by discursive relations of material inference-and-incompatibility. It is this line of thought that is then supposed to be completed by appeal to the features of such discursive practices that are made explicit by the normative and modal vocabularies that can be elaborated from those practices.

To fill in this last idea, I want to focus on the notion of <u>material incompatibility</u> that I have argued is implicit in discursive practice and which, each in its own way, both alethic modal and deontic normative vocabulary make explicit. As I have been telling the story, developing the meaning-use analytic pragmatist approach to semantics requires appeal to two different senses of 'incompatibility', which turn out to be related in a surprising

and revealing way. One is an objective *modal* sense: a matter of what states of affairs and properties of *objects* actually are incompatible with what others, in the world as it is independent of the attitudes of the knowing-and-acting subjects of practical, feedback-governed transactional engagements. If being made of pure copper is in this sense objectively incompatible with being an electrical insulator, then nothing can be both at the same time: it is *impossible* for one and the same object simultaneously to have both properties. That is a fact that holds regardless of how we use the *words* 'copper' and 'insulator'—indeed, it was a fact before there were any deployers of vocabulary at all. When, in the previous lecture, I showed how the concept of <u>incompatibility</u> could be used as the basis of a formal semantics capturing important features of the meanings of linguistic expressions, both logical and non-logical, this is the sense of 'incompatibility' that that semantic metavocabulary employed. The other sense of 'incompatible' is *normative*, and concerns *commitments* on the part of knowing-and-acting *subjects*—the ones who engage in discursive practices and exercise discursive abilities. To say that two commitments (whether doxastic or practical) are incompatible in this sense is to say that one cannot be *entitled* to both, and so that if one finds oneself with such commitments, one is *obliged* to *do* something: to rectify or repair the incompatibility, by relinquishing or modifying at least one of those commitments (to enter into a process of updating, of rectification, of further synthesizing a rational unity). What is incompatible with what in this sense is a matter of the practices and attitudes of the subjects of those commitments: the norms implicit in their behavior, what they in practice *take* or *treat* as incompatible in acknowledging and attributing the deontic statuses of commitment and entitlement.

The first point I want to emphasize is that these are clearly *different* notions of incompatibility. It is *impossible* for one and the same *object* to have incompatible *properties* at the same time. But it is merely *impermissible* for one and the same *subject* to have incompatible *commitments* at the same time. We *can* undertake such commitments. It is not impossible to do so. Indeed, we do it all the time—albeit usually involuntarily. When we do, the consequence is a change in *normative* status: we are not *entitled* to the incompatible commitments, and so are obliged to *do* something to rectify the situation. But we may not *actually* do what is in this normative sense demanded of us, or even practically be *able* to do it. We are discursively

born into a state of sin, and, for all our conscientious efforts, are by and large doomed to live in such a state. If p and q are incompatible in the *alethic modal* sense, then it is necessary that not (p and q). But if p and q are incompatible in the *normative deontic* sense, then it is indeed required that one not be *committed* to (p and q), in the sense that one *ought* not to be, but it does not at all follow that one *can*not be, or is *in fact* not so committed. The sort of looseness of fit between what is necessary or required in the deontic normative sense and what is possible or actual is not even intelligible in the alethic modal sense of 'necessity'.

It is worth noticing that these two senses of 'incompatible' are interdefinable with the two poles of the intentional nexus: knowing and acting *subjects* and the *objects* towards which their cognitive and practical states are directed. For (suppressing for present purposes the relativity to times[6]) objects are individuated by the way they 'repel' incompatible properties. It is *not* impossible for *two different* objects to have incompatible properties—say, being copper and electrically insulating. What *is* impossible is for *one and the same* object to do so. Objects play the conceptual functional role of *units of account for alethic modal incompatibilities*. A single object just *is* what cannot have incompatible properties (at the same time). That is, it is an essential individuating feature of the metaphysical categorical sortal metaconcept <u>object</u> that objects have the metaproperty of *modally* repelling incompatibilities. And, in a parallel fashion, subjects too are individuated by the way they normatively 'repel' incompatible commitments. It is not impermissible for *two different* subjects to have incompatible commitments—say, for me to take the coin to be copper and you to take it be an electrical insulator. What is impermissible is for *one and the same* subject to do so. Subjects play the conceptual functional role of *units of account for deontic normative incompatibilities*. That is, it is an essential individuating feature of the metaphysical categorical sortal metaconcept <u>subject</u> that subjects have the metaproperty of *normatively* repelling incompatibilities. A single subject just is what *ought* not to have incompatible commitments (at the same time).

These considerations show that although, as I have emphasized, the alethic and deontic senses of 'incompatible' are quite different, they are

---

[6] We may think of the time-references as built into the properties whose incompatibilities are being considered.

intimately related to one another. We are not faced with a term that is just ambiguous; the two uses of the word are not mere homonyms. Further, the relation between 'incompatibility' in the normative sense and 'incompatibility' in the modal sense is an expression of deep structural features of the nexus of intentionality: the nature of its *subjective* and *objective* poles and of the relation between them. What relates the two senses is a *process*, a practice, the exercise of an ability, a kind of practical *doing*: what discursive subjects are obliged to *do* when they find themselves acknowledging incompatible commitments—perhaps, as in the story about acid*, some acquired inferentially and some non-inferentially. *What* one is obliged to do is to *rectify* the incoherent commitments, by relinquishing one of the offending commitments, or, as in that example, modifying a mediating inferential commitment (and hence a concept). This updating is "repelling incompatibilities" in the normative sense. That objects "repel incompatibilities" in the modal sense is simply a fact: a *relational* fact metaphysically constitutive of objects as such. But subjects' repelling of incompatibilities is a *process*, an *activity*, a *practice*, the exercise of an *ability*. It is something they actively *do*. That they are *obliged* to do it is a fact metaphysically constitutive of subjects as such.

Here is the key point. By doing that, by engaging in the practice of rectifying commitments, subjects are at once *both* taking or treating the *commitments* involved as incompatible in the *normative* sense of obliging them to do something about that collision, *and* taking or treating two *states of affairs* regarding *objects* as incompatible in the *modal* sense that it is impossible for both to obtain. These are, I repeat, quite different senses of 'incompatible'. But in practically acknowledging an obligation to rectify or repair a set of commitments, one is doing something that can be specified not just by using one or the other, but, crucially, by using *both*. That it *can* be specified in both ways, both in normative terms and in modal terms, is what it is for the vocabulary whose use is being rectified to have semantic intentional *content*, for its deployment to count as *representing objects* and *saying of* them *that* they are objectively thus-and-so, for it to be the kind of *process* that establishes representational *relations*.

The basic phenomenon that underlies the complex meaning-use relations detailed in Figure 6.1 is accordingly the substructure shown in Figure 6.2.
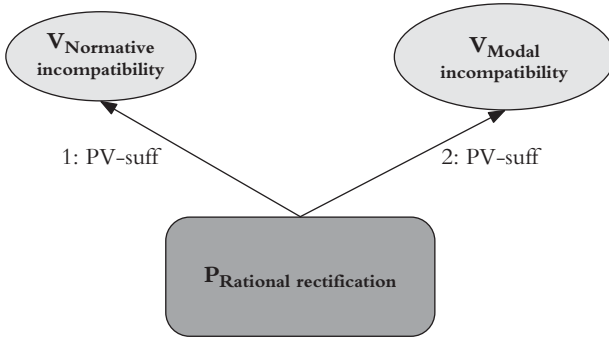
Figure 6.2 Rational rectification = treating commitments as normatively incompatible = treating properties as modally incompatible

Engaging in the practice of rectifying one's commitments is entering into a TOTE cycle whose exit condition is the removal of a local material incoherence in one's commitments, cognitive or practical. Practically acknowledging a commitment to that in some particular case is treating the commitments involved as discordant both in the sense that they are *normatively* incompatible for a *subject* and in the sense that they involve attributing *modally* incompatible properties to an *object*. We already saw that treating two (basic, non-quantificational) claims as incompatible is taking them to refer to one and the same object (as is taking them to be related as premise and conclusion of a material consequence relation). We are now in a position to see this fact as an aspect of a more general one. Shouldering the responsibility of repair and rectification of incompatible commitments is what one has to *do* in order to be taking one's claims to be *about* an objective world, in the normative sense of granting it *authority* over the *correctness* of one's claims. In treating two commitments as incompatible in the sense of *normatively* requiring giving up at least one of them (or otherwise modifying them so as to render them materially compatible) one is treating them both as claims *about* objects (about at least some of the *same* objects) and as *incorrect* about those objects because they attribute to them *modally* incompatible properties. That is what *subjects* must *do* in order thereby to be treating the vocabulary they deploy in acknowledging the commitments in question as expressing objective contents: claims about the properties (usually complex relational properties) of *objects*.

What we might think of as an *objectivist* semantic order of explanation begins with the way the world objectively is—construed here as a matter

of what really follows from what and what is really incompatible with what in the alethic modal sense. This conception is then taken to define the goal of inquiry, which accordingly determines a norm for the process of commitment-revision, according to an instrumental model. One ought to revise one's commitments so as more closely to approach the goal of practically *taking* commitments to stand in material incompatibility or consequence relations just in case the states of affairs and properties they represent objectively do stand in such relations to one another. This is what the subject is obliged, as a knower and an agent, to *try* to do. That ideal consilience of subjective *normative* attitude and practice, on the one hand, and objective *modal* fact, on the other, sets the standard for assessing the process of commitment revision. It is the "image of language triumphant" that process "draws within language militant," as Sellars puts it.[7] That never-reached but always-efficacious functional ideal is inscribed within the essentially *discursive* process of rectification and amplification by being the exit condition of a *practical* TOTE cycle that is mediated by relations of material incompatibility and consequence.

A complementary order of semantic explanation, by contrast, begins with what discursive practitioners actually *do*, that is, with the practical discursive process of rectifying and amplifying their commitments. It seeks to make the notion of objective modal relations intelligible in terms of this process, via pragmatically mediated semantic relations—relations of sense-dependence, not of reference-dependence, since the modal facts would be largely as they are whether or not anyone engaged in discursive practices. We have seen how normative talk of commitment and entitlement, and of the sort of incompatibility of commitments definable in terms of them, can serve as a pragmatic metavocabulary saying what subjects must *do* in order to be taking or treating two properties as objectively incompatible—as incompatible in the full-blooded modal sense that can then be understood as providing a second-order norm for assessing normative relations among commitments and entitlements. For that representational semantic normativity is implicit in the fact that rectifying one's commitments so as to eliminate *acknowledged* incompatibilities among them (discursive updating) is at once *both* treating two *commitments* as incompatible in the *normative* sense of obliging subjects

to *do* something, *and* treating two *properties* as incompatible in the objective *modal* sense.

So, in the most general terms, the objectivist order of semantic explanation proceeds from objective modal relations, via semantic normativity, to subjective normativity, and a complementary order of semantic explanation—what we might call "subjective pragmatism"—proceeds from the subjective normativity displayed in the practical activity of amplifying and rectifying acknowledged commitments, and seeks to understand in terms of that activity both objective modal relations and the semantic normativity (revealed as pragmatically mediated) that links them to subjective normativity.

The view I am recommending is inspired by the insights of what I have called "subjective pragmatism," but—as the symmetry of the meaning-use diagram I presented as expressing the complex resultant meaning-use relations between alethic modal and deontic normative vocabularies indicates—rejects its one-sidedness in favor of a more even-handed understanding. Rather than simply turning the objectivist order of explanation on its head, what I want to call "objective pragmatism" about intentionality sees those features of discursive practice that are made explicit by modal vocabulary and those that are made explicit by normative vocabulary as complementary, as each in principle fully intelligible only in terms of its relation to the other. Its understanding is, as the slogan that forms the title of this lecture has it, that discursive intentionality is a *pragmatically mediated semantic relation* that essentially involves both what one is *doing* in *saying* something, and what is *said* about how it is with what one is thereby talking *about*.

In this lecture I have principally been concerned to show how the practical activity of rectifying commitments by removing incompatibilities provides a perspective from which *deontic normative* and *alethic modal* vocabularies show up as two sides of one coin, as making explicit essentially complementary aspects of what then becomes visible as an intentional nexus semantically connecting knowing and acting *subjects* with the *objects* they know about and act on. This is the activity (the practice, the process) that *pragmatically mediates* the semantic relations characteristic of discursive intentionality. There is obviously a great deal more that could be said about the relations between the complementarity of these uses of normative and modal vocabulary and semantic intentionality, and about how to

understand semantic and representational *relations* in terms of the *processes* or practices that institute or establish them. My purpose here has been just to open up the topic in such a way as both to give some definiteness to the suggestion that discursive intentionality itself should be thought of as a pragmatically mediated semantic relation, to show how that relation can be understood as instituted by the pragmatic process of rectifying and amplifying commitments, and so to indicate some of the ways in which it may prove fruitful to think of intentionality in the terms of the analytic pragmatism that animates and is expressed in meaning–use analysis.

5   Conclusion

I took the overall title for this lecture series from an Italian proverb: "Between saying and doing, many a pair of shoes is worn out." I want to close with a brief reminder of some of the shoes I have been wearing out (perhaps along with your patience). I began with a picture of twentieth-century philosophical analysis that presents it as a tradition unified by a distinctive semantic project: systematically to explore different kinds of logical relations between the meanings expressed by different kinds of vocabularies. It is this semantic logicism that I see as distinguishing the various twentieth-century versions of the programs of empiricism and naturalism from their Enlightenment predecessors, and as making possible a third characteristic core program of analysis: functionalism in the philosophy of mind. My concerns in these lectures have been framed by the confrontation between this analytic semantic tradition and a kind of radical pragmatism, epitomized by (but by no means restricted to) the later Wittgenstein, which sees theories of *meaning* as in principle bad ways of thinking about the *use* of linguistic expressions—as resulting from intellectualist or scientistic misconceptions concerning the nature of discursive understanding. In its strongest form, the pragmatist challenge rejects the very possibility of general, systematic semantic analysis, leaving room only for a kind of local therapeutic pragmatics: diagnosis and treatment of misunderstandings of features of the use of particular expressions that are the residue of uncritical theoretical philosophical commitments.

But we need not choose between analyzing meaning and describing use. I have offered a number of suggestions as to how the insights of these two

traditions can be synthesized. One consists in showing how the vocabulary of automaton theory can be generalized to serve as a tool for analyzing practices-or-abilities, by specifying some as algorithmically decomposable into others, from which they can be algorithmically elaborated. I made two principal claims concerning this sort of purely pragmatic formal analysis. The first is that the fact that, as I claim, practices *sufficient* to deploy *logical* vocabulary can in this sense be algorithmically elaborated from practices *necessary* for the deployment of any autonomous vocabulary vindicates in a novel way the privileged position that logical vocabulary is accorded in the classical project of philosophical analysis. The second is that the pragmatic core of the artificial intelligence version of the program of functionalism should be understood to consist not in a thesis about the ultimately *symbolic* nature of sapience, but rather in the claim that autonomous discursive practices can be algorithmically elaborated from practices-or-abilities each of which can be engaged in or exhibited by non-discursive creatures. There are clearly many more kinds of important practical PP-sufficiency and -necessity relations. I pointed in particular to what I called "practical elaboration by training" (including "unsupervised" learning), and made a few suggestive remarks about issues that arise when we consider the combination of this with algorithmic elaboration, in the form of *pedagogical algorithms*.

Besides these kinds of purely pragmatic analyses of doings, I introduced the idea of a kind of analysis of the relations between saying and doing: what I called "meaning-use analysis." This is a way of representing and articulating the relations between meaningful vocabularies (along the way I've discussed logical, indexical, observational, modal, normative, and intentional vocabularies) and the practices-or-abilities of deploying them that constitute the *use* in virtue of which they *mean* what they do. To the sort of practical PP-sufficiency of one set of practices-or-abilities for another that obtains when one set can be algorithmically elaborated into another, meaning-use analysis adds two other important basic meaning-use relations: the PV-sufficiency of a set of practices-or-abilities to *deploy* a vocabulary, and the VP-sufficiency of a vocabulary to *specify* some set of practices-or-abilities. Composing these basic meaning-use relations allows us to exhibit more complex relations among vocabularies and practices (meanings and uses) as *resultants* of constellations of the basic ones. One simple but important example is the relation of one vocabulary being a *pragmatic metavocabulary* for another. It holds when one vocabulary is

VP-sufficient to specify practices that are in turn PV-sufficient to deploy another vocabulary. This is the simplest kind of *pragmatically mediated semantic relation*. It can happen that the expressive power of a pragmatic metavocabulary might differ substantially from that of the vocabulary it lets us say what we must do to deploy. This is the phenomenon I called "pragmatic expressive bootstrapping." For instance, we saw that it can be proven that automata PV-sufficient to deploy all recursively enumerable vocabularies can be specified in context-free vocabularies. And I argued that a non-indexical pragmatic metavocabulary can be VP-sufficient to specify practices PV-sufficient to deploy an indexical vocabulary.

In those two cases, the pragmatic metavocabulary is strictly expressively *weaker* than its target vocabulary. Sometimes the two simply have *different sorts* of expressive power. The big theme of the second part of my lectures has been the relations between *modal* and *normative* vocabularies, and of both to the autonomous discursive practices we attribute by the use of *intentional* vocabulary. In this connection I made five large claims:

- First, I argued that a suitably chosen normative (deontic) vocabulary (of <u>commitment</u> and <u>entitlement</u>) can serve as a sufficient pragmatic metavocabulary for alethic modal vocabulary. That is, we can explain what one must *do* in order to be deploying the objective modal notion of <u>incompatibility</u> in terms of the normatively specifiable notion of claimings that are incompatible just in case *commitment* to one rules out *entitlement* to the other.

- Second, I showed how the notion of <u>incompatibility</u> introduced that way can be used as the basis of a *semantic* metavocabulary, in terms of which we can define both *logical* and *modal* operators, and also represent important aspects of the contents of non-logical concepts.

- Third, I argued for what I called the "Kant-Sellars theses" (which constitute the first half of the complex resultant expressive property of vocabularies I called "universal LX-ness"): as is the case with logical vocabulary, practices-or-abilities PV-sufficient to deploy both alethic modal and deontic normative vocabulary can be algorithmically elaborated from practices PV-necessary for deploying any autonomous vocabulary.

- Fourth (which is the other half of the universal LX-ness of these vocabularies), I pointed out that normative and modal vocabularies

each make explicit (a matter of the VP-sufficiency of a vocabulary to specify practices-or-abilities) important aspects implicit in (PV-necessary for) the use of any autonomous vocabulary.

- Finally, as we have just seen, those complementary aspects made explicit by normative and modal vocabulary correspond to the subjective and objective poles of the intentional nexus between what discursive practitioners *do*, their activity of claiming, and the objects, properties, and facts that they thereby count as *saying* something *about*. This shows what is required for *practical* intentionality to develop into *discursive* intentionality. And it exhibits discursive intentionality as a particular kind of *pragmatically mediated semantic* relation.

I am going to close with a dark, but I hope intriguing suggestion. I think the view Hegel is trying to express with his notorious assertion of the identity of subject and substance[8] is that, conceptually, the normative sense of material incompatibility (his "determinate negation") that applies to subjective commitments, and the modal sense of material incompatibility that articulates objective facts and properties, are two sides of the same coin, each intelligible in principle only in terms of the other, because the *activity* of taking or treating two *commitments* to be incompatible in the subjective normative sense just *is* what it is to take or treat two properties or states of affairs as incompatible in the objective modal sense. I think this pragmatically mediated semantic relation (which has nothing to do with any claim about the *causal* dependence of how things are on how anyone takes them to be) is the essence of the view he develops under the heading of 'idealism'. But that is a story for quite another occasion.

---

[8] In the preface to the *Phenomenology* (1807, reprinted by Oxford University Press, 1979). Elsewhere (for instance, in the *Science of Logic* [1812–16, reprinted by Humanity Books, 1990]), he puts what I take to be the same point in terms of the identity of thought and being in what he calls the Idea.

Afterword: Philosophical Analysis and Analytic Philosophy

I

When I began my doctoral studies at Princeton in the early 1970s, Gilbert Harman was director of graduate studies. Among his characteristically idiosyncratic pieces of advice (along with "Start writing articles, like a grown-up, as soon as you can, and get out of graduate school as fast as you can—three years at the most") was the opinion that there is just no point in reading anything written more than five years ago: you want to find out about the current state of the discussion, and anything that hasn't been addressed in that length of time probably isn't important enough to bother with. The general view among my fellow students was that he represented the *ne plus ultra* of ahistorical approaches to philosophy. In time, I came to realize that they were dead wrong. At that stage of his career, Gil was steeped in the history of philosophy, and everything he wrote and thought was rooted in his understanding of it. The contrary impression resulted from failing to realize that he thought that, for practical purposes, philosophy had started with Quine.

I set up the story that I tell in the body of this work in terms of a reading of the classical project of philosophical analysis, as that project was developed during the twentieth century. This is a shorter historical run-up than I often require. In *Tales of the Mighty Dead*[1] I trace some

---

[1] Harvard University Press, 2002.

of the principal themes I develop in *Making It Explicit*[2] back to the early moderns. And these days I find it difficult to make any philosophical point without starting back with Kant and Hegel. In the case of the present project, though, I find that even some of those among my philosophical friends who are usually the most sympathetic to this way of proceeding question whether the enterprise I pursue here is really strengthened by framing it historically as a way of ''extending the project of analysis,'' as the title of my first lecture puts it. They point out, to begin with, that it need not be so construed. The usefulness and illumination provided by the metaconceptual apparatus I introduce for thinking about relations between meaning and use and the new sorts of pragmatically mediated semantic relations that apparatus brings into view—what it can teach us about discursive practices generally and the particular vocabularies I apply it to more particularly—does not at all depend on any continuity there might be between these methods and more traditional analytic ones. Given that independence, they argue, it is unwise to burden my project by associating it with what they see as a degenerating research program motivated by suspect methodological aspirations. When Richard Rorty first read drafts of these lectures, he asked, ''Why in the world would you want to extend the death throes of analytic philosophy by another decade or two?'' In a similar vein, John McDowell has described what I am doing here as perversely transplanting perfectly healthy pragmatist organs into the rotting corpse of analytic philosophy, so as artificially, and no doubt temporarily, to revive it as a kind of Frankenstein monster. I want to say something here about why, in the face of such strong reactions, I still want to endorse and defend not just meaning-use analysis and the idea of pragmatically mediated semantic relations, but also the further, optional characterization and commitment to these ideas as offering a potential way forward for what is recognizably a version of the classical project of philosophical analysis.

2

There are a number of different sources of discontent with that analytic project, and I think it is important to disentangle them. A significant

---

[2] Harvard University Press, 1994.

element of the enterprise pursued in these lectures is to respond to what I take to be the weightiest, deepest, and most important sort of objection to the classical project of philosophical analysis: the battery of considerations raised by the pragmatists, and above all Wittgenstein. But there are other sources of discomfort and lack of enthusiasm for the analytic enterprise, which I do not address in the body of this work. Some people do not see, even in retrospect, any good reasons why the advent of new logical tools should have ushered in a new philosophical era. Rorty once told me that it would not have mattered a bit to him, or to anything he had ever thought philosophically, if there had never been such a thing as formal semantics. (He delighted, in a Zen-masterish sort of way, in shocking me.) And I know others who maintain that nothing of real philosophical significance ever happened within three feet in any direction of a quantifier symbol. Such views are not only extreme, I think they are symptoms of a failure to appreciate some of the astonishing achievements and real promise of the logistical semantic tradition.

Here is an example that still fires my enthusiasm, after many years. In his seminal paper ''General Semantics,'' David Lewis invites us to pick whatever we like, depending on our more general philosophical proclivities, as semantic interpretants for declarative sentences and singular terms.[3] The semantically relevant whatsits associated with sentences might be, as Lewis prefers, sets of possible worlds, or as Michael Dummett prefers, sets of assertibility conditions, or something else. And singular terms might be interpreted by objects or by recognition conditions, or something else. Whatever we choose, it will be settled that the semantic intepretants of one-place predicates should be functions from the interpretants of singular terms to the interpretants of sentences. And it will be further settled that adverbs—which, like 'slowly', transform one predicate, 'walks', into another, 'walks slowly'—should be semantically associated with functions from functions from the interpretants of singular terms to the interpretants of sentences to functions from the interpretants of singular terms to the interpretants of sentences. When we then notice that their inferential behavior partitions adverbs into two classes, attributive and non-attributive, depending on whether it follows from one's having X'd ϕ-ly that one X'd—as one's having buttered the toast follows from having buttered

---

[3] The article first appeared in *Synthese*, 22 (1972), 18–67.

the toast in one's kitchen, but not from having buttered the toast in one's imagination—we are in a position to represent in set-theoretic terms *exactly* what the semantic difference between these two kinds of adverbs consists in. Getting this sort of algebraic grip on the *meanings* of non-logical expressions is a signal accomplishment. Even as we explore different approaches to semantics, we must be sure never to lose the precious comprehension it provides. It is, it seems to me, just the sort of thing that ought to inspire the philosophical imagination. At the least, I take it we are obliged to investigate just how far such methods can take us in making visible and formally tractable various other aspects of the content of the concepts we deploy in our philosophical reflections.

Another complaint that one sometimes hears is that analytic philosophy has come to focus on narrow, technical puzzles, scholastically generated by essentially self-contained literatures. An invidious contrast is then made between worrying about, say, the details of the behavior of proper names and indexicals in modal contexts, the pros and cons of four-dimensionalist views of spatio-temporal continuants, and the semantic paradoxes, on the one hand, and the best lessons to draw from the cataclysmic transition from traditional to modern culture, the essentially social character of self-consciousness, and the relative merits of art and science as revelatory of human nature, on the other. Which sets of sample concerns, it is asked, most deserves our allegiance as faithful to the august spirit of the tradition of *philosophia perennis*? No doubt there are a lot of toilers in the analytic groves who keep their noses very close to the ground indeed. But it does not follow, and it is not true, that, as a result, their work, for instance in the most technical reaches of the philosophy of language, relates to the philosophical tradition primarily as simply changing the subject. (Tempting as this conclusion may be. Bruce Kuklick describes the basically reactionary impulse of one of the great American philosophy departments of the first half of the twentieth century as presenting the dispiriting spectacle of "third-rate minds defending the rights of genius against the claims of technique."[4]) Providing disciplinary matrices within which the philosophical equivalent of normal science can proceed, between and alongside suggestions for paradigm shifts of various scopes, counts more in favor of the maturity of

---

[4] In "Philosophy at Yale in the Century after Darwin," *History of Philosophy Quarterly*, 21 (2004), 313–336.

a *Fach* than for its degeneracy. (This fact is just one reason that a *tu quoque* gesture at the most philological fringes of non-analytic philosophy does not amount to a constructive response to this challenge.)

In any case, although many philosophers working in the analytic tradition do not much concern themselves with how their concerns fit into a larger framework, I think we should think of the classical project of analysis as one way of working out a vision of us as essentially *discursive* beings, at once creators and creatures of our linguistic practices and the states of mind that make them possible and are made possible by them. This is the key into which traditional philosophical issues are transposed. In this sense, the linguistic turn characteristic of the twentieth century is not at all limited to the analytic tradition. One may, and should, take issue with the particular form in which that vision is worked out, but it is hardly to the point to complain that traditional issues of epistemology (and, more deeply, intentionality, agency, and normativity) have been left by the wayside because they are addressed in these terms. Even the narrower (hence more committive and controversial) versions of the analytic version of the linguistic turn, which identify it with the shift from the material to the formal mode (in Carnap's terms), that is, from concern with discursivity to specifically metalinguistic issues, ought to be understood as offering a way of working out the basically platonistic intellectualist strategy of understanding the implicit in terms of how it can be made explicit. And there have always been central analytic philosophers whose more detailed work took place in a larger, systematic philosophical context: Quine, Sellars, Davidson, and Dummett, to name just a few. (I'll return to this point below, when I consider the charge that analytic philosophy is committed to an outmoded and disagreeable kind of metaphysical project.)

A related (but I think distinguishable) complaint is that analytic philosophy has explicitly and (so) self-consciously cut itself off from the *history* of the discipline. There certainly has always been an ahistorical strand of thought in this tradition. Among the founders of the analytic tradition, Russell and Moore in particular promulgated an origin story that emphasized a radical break from all previous philosophy. The new logic was at last, for the first time, to put philosophy on the sure path of a science. But like other turn-of-the-century modernists in literature and the arts, the early analysts tended to overestimate the extent to which they broke with, rather than developed, the tradition they inherited and reacted

against. (Russell wrote and cared a lot about his hero Leibniz, and the continuities between Carnap's thought and that of his neo-Kantian teachers substantially outweigh the discontinuities, for instance.) As we achieve the temporal distance necessary for a dispassionate historical perspective on the history of early analytic philosophy itself, we ought to be able to take a more balanced view. In the body of this work I point out that the core programs of analytic empiricism and naturalism are recognizable descendants of their early modern forebears. And in *Tales of the Mighty Dead* I argue that this is true as well of the program of functionalism in the philosophy of mind. There, and in *Making It Explicit*, I also emphasize the strand of thought linking analytic philosophy to its past that Stanley Cavell epitomizes in the form of his characteristically trenchant aphorism: ''Kant depsychologized epistemology, Frege depsychologized logic, and Wittgenstein depsychologized psychology.''

It remains true that many later analytic philosophers continued to relegate the history of philosophy to the status of a minor sub-field, to the point of thinking of it as providing suitable employment only for those who, for one reason or another, were not capable of doing, or disposed to do, the real thing. I have already mentioned the widely shared suspicion that my teacher Harman once harbored such a view. It was his teacher, Burton Dreben, who expressed the attitude most forcefully, in his dictum that ''Garbage is garbage, but the history of garbage is scholarship.'' But this was certainly never a universally held opinion. One of the reasons Wilfrid Sellars is a particular hero of mine is the way he pursued analytic philosophy within a synoptic, systematic framework that was motivated and informed throughout by his intepretation of the history of philosophy, and above all of the significance of Kant's thought within that history. Sellars's avowed aspiration to move analytic philosophy from its Humean into its Kantian phase (and my own to move it from its incipient Kantian to a subsequent Hegelian phase) expresses a distinctive understanding of how the development of analytic philosophy fits into the larger context of the history of philosophy. Finding ourselves, as we do, in the midst of a golden age of analytic readings of Kant (initiated by Strawson and Bennett on the theoretical side, and Rawls on the practical), contemporary analytic philosophers should not be understood as bound by the propagandistic insistence that all things have been made new in analysis that was put forward by some of the founding members as an integral part of the fighting faith

they crafted as a crucial weapon in the struggle to make disciplinary room for their nascent movement. (On one point, however, they might be proven to have been prescient. Recoiling from the peculiar form of Hegelianism developed by their British Absolute Idealist teachers, Russell and Moore advocated a particularly radical surgical intervention to prevent the spread of the infection to which they themselves had earlier fallen prey. They took it that the idealist rot had set in already with Kant, who accordingly must be excised along with Hegel from the expressively progressive tradition they were retrospectively reconstructing—hence Russell's ''Zurück nach Leibniz'' campaign. They were confident that one could not open the door far enough to let Kant slip through and then close it quickly enough to keep Hegel out. And on this point of tactics—whatever the advisability of their overall strategic gatekeeping enterprise—they may well turn out to have been correct. We'll see.)

3

A fourth objection to the whole enterprise of analytic philosophy is that one feature that binds together its various, otherwise disparate, phases, schools, and projects is a commitment to an objectionable *scientism*. I think there is indeed a telling objection that belongs under this heading, but it is important to be clear about what it is, and to distinguish it from near relatives that in one way or another miss the mark. For there are many things to mean by 'scientism'. One broad category is what might be called *Enlightenment* scientism. It is a valorization of scientific activity that privileges it as the paradigmatic, highest, and most revealing expression of our humanity—by contrast to the Romantic aestheticism that defined itself by according that privilege instead to artistic activity. (Of course, we students of that great synthesizer, the Romantic rationalist Hegel, would like to think that we have the conceptual resources to evade any necessity simply to make a choice here. But that aspiration is not my current topic.) Analytic philosophy has historically endorsed a variety of views that count as scientist in this general sense. I think the most important issue in the vicinity concerns how one thinks about the relations between two forms of understanding, and I'll discuss that presently. But that issue is not raised

by at least one very thin form that analytic philosophy's identification with the scientific enterprise has taken.

What might be called '*sociological* scientism' is largely a matter of style. It is the idea that the shape of the philosophical community, and of its professional discourse, should be modeled on that of natural scientists—rather than that of, say, café intellectuals, political activists, literary critics, novelists, or poets. Philosophy is thought of as professionalized, as a discipline, as instituting, through shared training, common standards for assessing arguments and accomplishments. It is thought of as a communal, co-operative, cumulative undertaking, in which progress is made by building on the agreed-upon achievements of others, and as a *cognitive* enterprise in the sense that it aims to produce and extend a publicly available kind of *knowledge*, or more broadly, *understanding*. Relations among its practitioners are accordingly to be respectful and collegial, personal differences and ambitions are to be submerged in the pursuit of the larger common, impersonal goal. This ideal, which in the eyes of some of its most ardent avatars in the Vienna Circle was closely associated with a progressive social and political agenda, has, I think, served philosophy rather well.[5] Though its virtues are not free of associated vices, its scientism about philosophy in this sociological sense does not seem to me to provide any reason to reject the classical project of philosophical analysis.

More substantively, 'scientism' can just mean '*naturalism*'. It would be wrong to identify that core program of analysis with the whole analytic project, however. Not all analytic philosophers are philosophical naturalists. Moore certainly was not. Some logical empiricists were not: the Schlick/Neurath split within the Vienna Circle revolved around the issue of which of empiricism or naturalism should yield when the two programs collide (as they did, for instance, over the status of alethic modality). And contemporary analytic consciousness studies—carrying on one strand of the empiricist tradition—very often is not only not pursued in a naturalistic key, but offers a distinctive set of *anti*-naturalist arguments. Furthermore, philosophical naturalism about topics (vocabularies) other than philosophy itself need not be thought of as a form of scientistism about philosophy.

---

[5] An interesting perspective on this issue is provided by Jacques Bouveresse, in his essay ''Reading Rorty: Pragmatism and its Consequences,'' in Robert Brandom (ed.), *Rorty and His Critics* (Blackwell, 2000), 129–45.

Here Sellars's *scientia mensura* dictum may serve as a paradigm: "In the dimension of describing and explaining the world, science is the measure of all things, of what is, that it is, and of what is not, that it is not."[6] But Sellars does not take it that normative vocabulary, for instance, is in the business of describing or explaining. For that reason, he does not take it that intentional vocabulary is either. And he takes the task characteristic of philosophy precisely to be offering an account of the relations between such vocabularies and the vocabularies whose job *is* "describing and explaining the world," for whose use he takes it that natural science is authoritative. (In this respect he accordingly fits the definition of the classical project of analysis I offer in my first lecture.)

I think the charge of scientism begins to get a real grip when the admiration of science characteristic of Enlightenment scientism moves beyond its merely sociological and naturalistic forms and manifests itself as *methodological monism*. This is the view, roughly, that scientific knowledge is *the* form of knowledge, and scientific understanding is the only kind of understanding that deserves the name. This line of thought gives a very strong reading to the Kantian project of "putting philosophy on the sure path of a science." Insofar as philosophy is a cognitive enterprise at all, it must be continuous with the empirical, natural sciences—destined eventually to take its place as part of the single Unified Science. (One would hardly hold this view if one were not a naturalist about many other vocabularies apart from philosophy. But the point I just tried to make is that that kind of naturalism in no way commits one to this further step.) On this view it is a profound and dangerous mistake to think that the difference in their subject matters brings with it a substantive and fundamental difference in method and aim between the *Naturwissenschaften* and the *Geisteswissenschaften*. In particular, discursive practices and practitioners, vocabularies and the concepts deployed in using them, do not in principle require methods of investigation, or support kinds of understanding, different from those that have proven so successful in addressing the non-discursive world. This methodological commitment ties together representative figures of the analytic tradition otherwise as diverse as Russell, Ramsey, Carnap, Quine, and Fodor.

---

[6] §42 of *Empiricism and the Philosophy of Mind*, edited and with a Study Guide by Robert Brandom (Harvard University Press, 1997).

Now, I reject scientism in the form of methodological monism. Here again, care is called for, to be clear about the reasons that warrant such rejection, and about just what follows from it. In spite of its popularity among analytic philosophers, I do not think a commitment to methodological monism needs to be understood as an integral part of the analytic project itself. But neither is the association simply accidental. There are considerations that *are* integral to that project that can easily appear to call for this methodological stance. Some distinctions are in order. One way into the issue is provided by a sort of *modus tollens*. One constant in Wittgenstein's thought, early and late, is his denial of methodologically monistic scientism. ''Philosophy is not one of the natural sciences,'' he says in the *Tractatus*,[7] and this view seems to be part of what lies behind the theoretical quietism of the later work. In fact, I think Wittgenstein thinks that if systematic philosophical theorizing *were* possible, it would mean that philosophy *is* an empirical science. Since it is not, philosophers must eschew theorizing, restricting themselves instead to light, local descriptions of discursive practices, where such descriptions might provide helpful reminders in freeing ourselves from the sorts of misunderstandings and puzzlements that arise precisely from the theories implicit in inherited pictures of what is going on when we think and talk. Whether or not Wittgenstein himself reasoned in this way, I take it that it is common for his admirers to see him as presenting us with a forced choice: either embrace scientism about philosophy of the methodologically monistic sort—that is, take philosophy to be an empirical, scientific discipline—or give up the idea of systematic philosophical theorizing once and for all.

I think this is a false choice. Rejecting scientism of the methodological monistic sort does not entail giving up the possibility of systematic philosophical theorizing about discursive practice. One of the most powerful methodological features of the natural sciences is the postulation of unobservable theoretical entities, and their deployment in constructions aimed at explaining what is observable. Theoretical entities are those about which we can make only theoretical, and not observational, claims. Theoretical claims are ones that we can only become entitled to as the conclusions of *inferences* from other claims, not non-inferentially, as the results of exercising reliable dispositions to respond differentially to environing states of affairs

---

[7] Proposition 4.111.

by making observation reports of them. A generalization of this method would have the role played by observational vocabulary played by any antecedently available vocabulary, whether observational or not. So for instance one might postulate *meanings* to explain proprieties of *use*, where the latter are expressed in a non-semantic vocabulary, whether or not our access to claims about correct usage are made observationally or themselves inferentially.[8] The claim that theorizing of this sort could be legitimate in philosophy does *not* commit one to the claim that this method is the *only* legitimate method of acquiring philosophical understanding—which is what methodologically monistic scientism claims. The generalized method of postulation and construction might be *one* form of philosophical understanding among others. I want to claim that what is objectionable about the methodologically monistic form of scientism is its exclusivity. Rejecting that at least leaves open the question of whether, and which, features of natural scientific investigation, explanation, knowledge, and understanding ought also to be counted among those useful and appropriate in philosophy. After all, *description* is also a central and essential element of scientific methodology, and even the most rigorous versions of Wittgensteinian quietism allow philosophers to describe features of our linguistic practice.

4

Asking this question, and seeing that it is not settled by rejecting methodologically monistic scientism, then raises a fifth sort of objection to the project of philosophical analysis, one that is a somewhat more careful and nuanced successor to the previous one. This is that the defining aim of philosophical analysis is to achieve a kind of understanding of vocabularies and discursive practices-or-abilities to use or deploy vocabularies that is in principle not available, because that subject matter requires *another*, quite different sort of understanding. The kind of understanding sought by analytic philosophers is appropriate only to non-discursive activities. The reason we should not be methodological monists is that understanding

---

[8] The analogous postulation of intentional states to explain behavior Sellars calls "philosophical behaviorism," by contrast to the "logical behaviorism" that is committed to *defining* the states in terms of behavior. In the case of meaning and use, the corresponding non-theoretical move is a semantic instrumentalism that insists, as Dummett used to do, that every aspect of meaning be *manifestable* in use.

talking and thinking, concept use, vocabularies, natural language utterances and texts, is a distinctive sort of achievement. This kind of understanding, what we might call *hermeneutic* understanding, is not expressible in explicit rules, formalizable in regimented technical or artificial languages, as analytic philosophers are committed to doing by the semantic logicism characteristic of the classical project of analysis. The mathematized mature natural sciences have had great success in achieving what we might call *algebraic* understanding of great swathes of the inanimate natural world. (Whether the animate biological world, including sentient-but-not-sapient creatures and their activities, itself already calls for further special sorts of understanding remains a lively and controverted question.[9]) But when the topic is *culture* rather than *nature*, another sort of approach is called for. Here the paradigm of understanding is that exhibited by competent native speakers of natural languages when confronted by everyday utterances expressed in familiar vocabulary. This sort of practical grasp of meanings (the medium of the cultural) is not in the most fundamental cases a matter of explicit theorizing at all. And it is not a matter of mapping or translating the utterance into some other vocabulary (perhaps with the use of auxiliary logical vocabulary) either. (In the sense that matters for this point, the language of my thought is just my language: a language I speak.) More sophisticated forms of hermeneutic understanding, of the sort exercised by the literary critic, jurisprudential interpreter, and reader of philosophical texts, are possible, but they are both rooted in the basic one and do not come closer to having the structure of algebraic understanding.

A pragmatist line of thought common to the Dewey of *Experience and Nature* and *Art and Experience*, the Heidegger of *Being and Time*, and the Wittgenstein of the *Philosophical Investigations* is that there is such a thing as hermeneutic understanding in this sense, it is a genuine and distinctive kind of understanding, and it is the most basic kind of understanding, in the sense that all other sorts of understanding are parasitic on it and develop out of it. It is the primordial sort of practical discursive know-how: the capacity to engage in an autonomous discursive practice. In particular, they are concerned to insist that the sort of algebraic understanding characteristic of mature mathematized sciences—the sort for which analytic philosophers

---

[9] See the discussion in Michael Thompson's astonishing, original *Life and Action*, forthcoming from Harvard University Press.

long—is pragmatically dependent (PP-dependent) on everyday hermeneutic understanding, which accordingly cannot be replaced by, or reduced to, the more technical kind.

I accept all of these pragmatist claims about the distinctiveness and basicness of ordinary hermeneutic understanding of discursive performances and their products. Should we conclude that the analytic project is just a mistake? I don't think so. For this pragmatist line of thought does not entail that many aspects of discursive practice might not *also* be susceptible to understanding of the sort I have called "algebraic." And where it is possible, broadly algebraic understanding has distinctive virtues, which adherents of the project of analysis are right to esteem and treasure. The slogan of the analytic project is "Faith, hope, and clarity—and the greatest of these is clarity." The clarity in question is specifically *conceptual* clarity. It would seem to have two dimensions: definiteness and perspicuity. From a pragmatic point of view, the significance of a speech act is definite insofar as its normative significance is settled. From the point of view of semantic inferentialism, this means that concepts are definite insofar as their circumstances and consequences of application are settled: when one is committed and entitled to apply them, and what such application commits and entitles one to. Perspicuity is then epistemic or psychological ease of access to those circumstances and consequences of application. On this line, thinking clearly is both formulating one's claims (claimables) so as to fix what one would be committing oneself to by endorsing them and what would entitle one to do so, and being aware of those definite consequences and circumstances of application of the concepts that articulate the contents of the concepts one is applying. Writing clearly is choosing one's words so as both to determine the inferential boundaries (or, one could equally well say, truth conditions) of one's claims and to convey them to the reader.

What I'm calling the "algebraic" form of understanding achieves clarity along both the dimensions of definiteness and perspicuity by *constructing* the conceptual contents expressed by a target vocabulary. It does that by exhibiting them as complexes formed as the products of applying explicit algorithms to the conceptual contents expressed by a base vocabulary (treated for this purpose and relative to this construction, as simple). From the broadened perspective made available by thinking about pragmatically mediated semantic relations, algorithmic elaboration of one set of vocabulary-deploying practices-or-abilities into another is, as I point out in

Lecture 2, a generalization of the traditional semantic logicism characteristic of the classical project of philosophical analysis. What corresponds in this semantic-analytic project to the postulation of unobservables in empirical scientific theorizing is the employment in the algorithmic construction also of some further auxiliary vocabulary, whose use is not governed by antecedent norms but is determined instead by stipulated inferential connections to both base and target vocabulary.[10] This algorithmic-constructional method (building complex things by applying well-defined operations to simpler things) is a very good, perhaps superlative, way of securing clarity of understanding. I have elsewhere called it for this reason the "gold standard" of understanding generally—by which I mean that *when and to the extent it is available*, it is the very best sort of understanding to have, for it takes the issue of what one means (what one is committed to by a claim, what is incompatible with it, what would count as evidence for or against it, and so on) out of the hands, out from under the authority, of the one making the claims. It establishes a fact of the matter about the inferential relations that articulate the contents of the concepts expressed by the target vocabulary that swings free of the beliefs and preferences of the concept user: what she would *like* to be committed to or treat as evidence for those claims. If a dispute arises, those who are assessing the claim in question can say, with Leibniz, "Let us calculate." This aspiration to develop "a general method in which all truths of reason would be reduced to a kind of calculation … and errors—except those of fact—would be mere mistakes in calculation"[11] (I take it that "truths of reason" here stands in for inferential relations that articulate the contents of the concepts involved) is one of the reasons Leibniz was a hero for Russell in the latter's attempt to develop a notion of philosophical analysis. This sort of clarity of understanding is a pearl without price—all the more to be prized where the target vocabulary it concerns is weightier and more difficult, as is the case with many of those

[10] Thought of in this framework, in the case of empirical scientific theorizing, the base claims and the target claims are formulated in the same antecedent vocabulary—which may be, and must include, observational claims in the strict sense of those elicited by the exercise of reliable noninferential differential responsive dispositions, but which also include statements couched in the vocabulary (including theoretical, that is, only inferentially applicable vocabulary) of other scientific disciplines, for instance, those that address the workings of measuring instruments and the ranges of counterfactual robustness of various collateral premises.

[11] In his 1685 ms. "Machina arithmetica in qua non additio tantum et subtractio sed et multiplicatio nullo, divisio vero paene nullo animi labore peragantur," Mark Kormes (trans.), in D. E. Smith (ed.), *A Source Book in Mathematics* (McGraw-Hill, 1929), 173–81.

either used or addressed by philosophers. (This sort of clarity facilitates communication—scientism in the sociological sense.)

Appreciating this cardinal virtue of the algebraic form of understanding does not require taking issue with the pragmatist point that it is in principle parasitic on and intelligible in principle only against the background of a more basic sort of practical discursive understanding that does not at all have this explicit theoretical form. It is useless—for instance, in settling disputes about what someone is committed to by a claim couched in the target vocabulary being (re)constructed—unless there is a shared base vocabulary about whose proper use all parties can agree in their practice. We are not in a position to calculate unless we can all practically go on in the same way in counting and adding—as Wittgenstein is at pains to remind us in many different ways and many different contexts. And the same is true of algebraically computing the inferential roles or truth conditions of complex expressions from those of simpler ones. Algorithmic elaboration is a way of *leveraging* practical agreement in the use of one vocabulary into practical agreement in the use of another. It is true that what plays the role of a base vocabulary for one such constructive enterprise may be the target vocabulary whose proper use is algorithmically reconstructed by another. But the point Wittgenstein was after here is that it cannot be algorithmic elaboration all the way down. At some point each such chain must be anchored in practical agreement about what it is and is not correct to do with a vocabulary that is *not* settled by being algorithmically handed off to some prior one. And that is to say that we should *not* make the jump from the legitimate *local* aspiration to be able to settle some semantic-inferential disputes in the "Calculemus" way to Leibniz's dream of a *global lingua characteristica, all* of whose concepts are governed by a *calculus ratiocinator* that is in this sense *universal*.

Acknowledging the value of the unique clarity afforded by algebraic understanding accordingly does not entail commitment to this sort of understanding being available in every case, even in principle. It does not oblige one to embrace the shaky method of the drunk who looks for his keys under the streetlamp, not because they are likely to be there, but just because the light is better there. We should admit that, sometimes, algebraic understanding is not available—indeed, that *every* context in which it *is* available contains an appeal to a base vocabulary whose use is *not* held in place algebraically, but depends on *another* sort of practical mastery

and understanding. Algebraic understanding can be no more legitimate than the hermeneutic understanding on which it depends and which it leverages, amplifies, and concentrates. It follows that philosophy cannot be *identified with* analysis, thought of as comprising the tasks of understanding algebraic understanding and applying it in semantics. Even under the broad heading of trying to understand discursive practice, there is a more basic sort of hermeneutic understanding, whose implicit, practical, everyday species and whose explicit, theoretical, sophisticated species must both be studied and exercised by philosophers. Thinking through the presuppositions of its project shows that analytic philosophy can aspire at most to being one species of the genus. (In the third, methodological, chapter of *Tales of the Mighty Dead*, and again in "Hermeneutic Practice and Theories of Meaning"[12] I try to say something specific and systematic about how the different aspects of discourse addressed by these two sorts of understanding and their associated disciplines complement one another.)

5

There is, then, a lot more to be understood about discursiveness than can be understood algebraically. This is obviously true *de facto*, and I have just rehearsed an argument that it is true also *de jure*. But can we know *in advance* that the algebraic sort of understanding is not available at all for some subject matters? Might it not be the case that the very nature of discursive practice makes it unsuitable for this sort of account? Perhaps philosophical analysis must inevitably "murder to dissect," the very method it employs making it impossible for it ever to grasp the essence of the phenomenon it addresses. The image reminds us that claims of this same general form concerning not discourse, but life, were a commonplace in the early nineteenth century, as biology broadly shifted its concern from structure to function, from anatomy to physiology. Hegel takes over the Romantic trope of organic unity, and in a characteristic and central move transposes it into a rationalist key, by treating the sort of unity of disparate elements characteristic of organisms as an image of the semantic or conceptual *holism* he had discovered to be a central structure of the discursive—as the way nature

---

[12] *SATS: Nordic Journal of Philosophy*, 5/1 (2004).

already implicitly expresses a fundamental structure that becomes explicit in culture. Perhaps the holistic character of meanings or conceptual contents, of thinkables and sayables, in principle resists reconstruction by bottom-up, compositional, ultimately atomistic algebraic elaboration? Maybe by its very nature it demands instead a top-down, holistic hermeneutic form of understanding.[13]

I think this is a substantive challenge. It demands a much fuller discussion than I can enter into here. The conditional ''if discourse is semantically holistic, then in principle no algebraic reconstruction of conceptual contents or meanings is possible,'' is one that can be exploited either by *modus ponens* or by *modus tollens*. One sort of critic of analysis, inspired by Hegel, adopts the former strategy. Analytic semantic theorists such as Jerry Fodor adopt the latter. My own strategy is to dispute the inference codified in the conditional. One of the satisfying results of the incompatibility semantics sketched in my fifth lecture is the algebraic construction of a counterexample to some of the most powerful arguments for that conditional that have been offered from the analytic side. And in the sixth chapter of *Tales of the Mighty Dead* I consider analytically how one might get around some of the most worrisome difficulties that I think Hegel, the first philosopher to think hard about the issue of semantic holism, already wrestled with. But I think it is early days in our discussion—that we have only begun seriously to come to grips with this knotty issue. I suspect that it will resist Gordian solutions, yielding only gradually to the application of insight, patience, and analytic ingenuity. We have as yet, I think, no useful algebraic representation of the seriously multi-premise, non-monotonic, material inferences that are responsible for the phenomenon of semantic holism.

Putting aside the particular issue of semantic holism, it is at any rate important to keep in mind that the claim that there are some vocabularies, some discursive practices-or-abilities, that are by their very nature not amenable to analytic algebraic reconstruction does not *follow* just from the observation made above (in denying methodologically monistic scientism)

---

[13] Notice that in Lecture 1 I explicitly confront a corresponding challenge to the in-principle possibility of analysis, associated there with Wittgensteinian pragmatism, that appeals to the dynamic developmental-historical character of vocabulary use, rather than to its holistic aspects. Hegel, of course, was concerned with both, and, as I read him, with developing a notion of determinate conceptual content that would exhibit them as two sides of the same coin.

that every analysis or algebraic reconstruction of a target vocabulary must make use of and so depend on the prior semantic determinateness and understanding of what is expressed by some base vocabulary. That is, it does not follow that there is some order of, as it were *natural* basicness among vocabularies, which must have unexplained unexplainers (base vocabularies that do not admit of analytic algebraic reconstruction in terms of others) as its most basic elements. It might well be that although each analytic–algebraic account of the use of any vocabulary must appeal to some base vocabulary whose use is not explicated in that account, every vocabulary that plays that role of base vocabulary in some analyses plays the role of target vocabulary in some other successful analysis. A claim of the form $\forall x \exists y[Rxy]$ does not entail one of the form $\exists y \forall x[Rxy]$. (It is true that the world has a population problem because during every minute there is a woman somewhere in the world having a baby. But it is not a productive way to address the problem to look for the woman who is having all those babies and make her stop doing what she is doing.) The sense in which algebraic understanding rests *de jure* on hermeneutic understanding may be merely of the local, $\forall x \exists y$ sort, not the global, $\exists y \forall x$ sort.[14]

Distinguishing between taking the base-vocabulary/target-vocabulary distinction, in terms of which I have characterized the classical project of philosophical analysis, to be local and relative to particular expressive-explanatory undertakings, on the one hand, and taking it to be global and absolute, on the other—between taking it to be a matter of cognitive convenience or taste, and taking it to be something we could get substantively wrong because of how things anyway are—brings into view the notion of <u>universal base vocabularies</u>. That notion is at the center of the sixth and final sort of objection to the whole analytic philosophical enterprise that I will consider here. That objection arises first in an *ad hominem* form, as depending on subsequent commitments I undertake, collateral to the strategy of framing my presentation of meaning-use analysis as a way of extending the classical project of analysis by taking into consideration pragmatically mediated semantic relations. But the concerns it raises are deep and general,

---

[14] Picking up an earlier remark: This is admittedly a small and elementary philosophical point. But it guards us against what can be a tempting substantive mistake in thinking about the implicit practical background of the explicitly theorized. And making it clearly requires getting within three feet of the quantifier symbols.

and their significance extends well beyond that parochial setting. The *ad hominem* point is the observation that although I take the classical project of philosophical analysis to be worth extending and developing, I reject both the core project of empiricism and the core program of naturalism. (In my third lecture I present reasons to be skeptical also about the core program of AI-functionalism—an issue to which I shall return below.) One might then ask what interest the classical project of philosophical analysis retains when shorn of those programs (and any successors that share their objectionable features), and whether the same considerations that render those programs objectionable accordingly impugn the more general umbrella project that sheltered and nurtured them. These questions turn on the issue of what is wrong with the empiricist and naturalist programs. The complaint I am concerned to assess diagnoses them as *metaphysical* programs, and urges us to reject them for that reason. And it claims that the analytic project generally is of philosophical interest only insofar as it takes the form of a program predicated upon such disreputable and ultimately insupportable metaphysical commitments.

What is distinctive of empiricism and naturalism, considered abstractly, is that they each see some one vocabulary (or vocabulary-kind) as uniquely privileged with respect to all other vocabularies. Empiricism takes its favored vocabulary (whether it be phenomenal, secondary-quality, or observational) to be *epistemologically* privileged relative to all the rest. In what I think of as its most sophisticated forms, the privilege is understood more fundamentally to be *semantic*, and only derivatively and consequentially epistemological. Naturalism takes its favored vocabulary (whether it be that of fundamental physics, the special sciences, or just descriptive) to be *ontologically* privileged relative to all the rest. In both cases, what motivates and gives weight and significance to the question of whether, to what extent, and how a given target vocabulary can be logically or algorithmically elaborated from the favored base vocabulary is the philosophical argument for epistemologically, semantically, or ontologically privileging that base vocabulary. These are arguments to the effect that everything that can be known, said, or thought, every fact, must in principle be expressible in the base vocabulary in question. It is in this sense (epistemological, semantic, or ontological) a *universal* vocabulary. What it can*not* express is fatally defective: unknowable, unintelligible, or unreal. One clear thing to mean by 'metaphysics' is the making of claims of this sort about the universal expressive power of some vocabulary.

6

The general objection can now be put in the form of two claims:

1. The philosophical interest of the analytic project depends on endorsing programs (paradigmatically empiricism and naturalism) that are metaphysical in this sense.
2. Metaphysics in this sense is a bad idea.

The conclusion is that we should reject the classical project of philosophical analysis root and branch.

An argument for the first claim might go like this. Why would we care about showing that what is expressed in some target vocabulary can or cannot be expressed in a base vocabulary (when suitably elaborated), unless we thought for independent and antecedent reasons that, by doing that, at least in some cases we were settling what could and could not be known, what is and is not intelligible, what is and is not real (or some other philosophical issue of that magnitude)? Apart from payoffs such as those, why not just acknowledge that every vocabulary is the one it is, and not some other one? Different vocabularies have different expressive powers. So what? Taking the expressive powers of *some* vocabularies to be authoritative for settling issues of great pith and moment for epistemology, semantics, or ontology is the recipe for producing a responsive answer to the "so what?" question. Doing that is taking those vocabularies to be *universally* privileged base vocabularies, with respect to some large-scale philosophically significant assessment. And that, according to the definition we are working with, is undertaking a distinctively *metaphysical* commitment. In fact, the history of philosophy in the twentieth century offers some empirical confirmation for this conceptual point. The first wave of analytic results was logicist in a strict sense: not only the means of semantic elaboration, but also the base vocabulary itself was *logical* vocabulary. (Here I have in mind Frege's *Begriffsschrift* and *Grundgesetze*, Russell's "On Denoting" and *Principia Mathematica*, and Wittgenstein's *Tractatus*.) But after that, beginning already with the Cambridge analysts of the 1920s, and accelerating with Ayer and the Vienna Circle in the subsequent decades, and continuing well past mid-century for instance with Quine, the form in which philosophical analysis really took hold was indeed in connection with what for that very reason I call the "core programs" of empiricism and

naturalism. Of course, there is a delicious irony here—manifest to those working outside this philosophical tradition, but often invisible to those within it. For one important, perhaps dominant, current of classical analytic philosophy defined itself by contrast to its predecessors and contemporary rivals precisely by its *anti*-metaphysical tenor and fervor. Granted, Bertrand Russell, Peter Strawson, and David Lewis would not have been surprised or upset to be accused of engaging in metaphysical undertakings. But Carnap, Ayer, Quine, and Hempel certainly would.

The second limb of the objection is that metaphysics in this sense is in principle a defective enterprise. I am not really sure how arguments for such a claim at this level of generality go. I have heard four sorts, and am somewhat moved by a fifth. Empiricists reject metaphysical claims because they want to make claims there can be empirical evidence for, because they take that to be a necessary condition of those claims being candidates for expressing *knowledge*, or, indeed, *meaning* anything. Of course, that this methodological principle contradicts their own empiricist metaphysical principles, including this one, was forcefully pointed out by Hempel in his masterful "Empiricist Criteria of Cognitive Significance."[15] Corresponding considerations, of course, show that one who rejects empiricism on the grounds that it is unacceptably metaphysical cannot offer these empiricist reasons for rejecting metaphysics. Naturalists often reject metaphysics on the grounds that it is not a product of natural science. When we have real physics, why should we want, and how would we be justified in adding, *meta*physics? From this point of view it seems like the attempt to add an otiose layer of *hyper*-physics. Or, it is an attitude toward the results of science that is itself not science, but a kind of scientism that itself has no scientific credentials. Arthur Fine's rejection of scientific realism in favor of the "natural ontological attitude" is a sophisticated version of this thought.[16] Again, though, someone who rejects naturalism as objectionable metaphysics cannot do so for these naturalistic reasons.

Pragmatists such as Rorty object to the privileging of some vocabularies as universal base vocabularies characteristic of metaphysics on the grounds that it depends on a false, indeed, ultimately magical, understanding of

---

[15] First published in 1950; reprinted in *Aspects of Scientific Explanation* (Free Press, 1970).
[16] "The Natural Ontological Attitude," in D. Papineau (ed.), *The Philosophy of Science* (Oxford University Press, 1996).

the nature of the sorts of privilege or authority involved. All normative statuses, including these, are instituted by social practice. There are no such normative statuses apart from our practical normative attitudes of taking or treating some things as privileged or authoritative. And for instrumentalist pragmatists like Rorty's hero Dewey (whom he *sometimes* follows in this respect), the ultimate source of those attitudes is our own needs, wants, and convenience. What is wrong with the metaphysical sort of privileging of vocabularies is that it requires the idea of some vocabulary being necessarily privileged by *how things are*—God's vocabulary, or Nature's, or even Mind's, or Meaning's vocabulary—quite apart from our contingent projects and attitudes. Once again there would seem to be an issue about the self-referential stability of this view: is the social nature of normativity, and its normative capacity to trump metaphysical projects, just a feature of how things anyway are? Rorty's response is that this commitment, too, is indeed to be assessed relative to our needs, welfare, and projects. We learned (well, we were supposed to have all learned, though current events make it dubious that we did) from the Enlightenment that it was bad for us in our development as mature humans in charge of our lives and institutions, to understand *moral* normativity as simply reflecting how things were with a non-human (albeit divine) reality. And Rorty's practical proposal for a second Enlightenment, completing the work of the first, is to extend that lesson of self-reliance from the practical to the theoretical sphere, for reasons analogous to those that warrant the first move.[17] This is radical and controversial.

A more Wittgensteinian pragmatism addresses metaphysical programs more in a retail than a wholesale spirit. It addresses empiricism and naturalism, and any successor projects one by one, seeking to undermine the specific claims of privilege they put forward. (Here the various criticisms Sellars addresses to empiricism, adverted to in my opening lecture, as well as those of Quine and Austin can serve as paradigms.) But it also expresses a more general suspicion that any such program will turn out, upon examination, to have been motivated by a philosophical anxiety that can be

[17] See, for instance, Richard Rorty, "Pragmatism as Anti-authoritarianism," *Revue Internationale de Philosophie*, 53/1 (1999): 7–20, and "Universality and Truth," in Robert Brandom (ed.), *Rorty and His Critics* (Blackwell 2000). I offer an assessment of this argument in section III of "When Philosophy Paints its Blue on Gray: Irony and the Pragmatist Enlightenment," *boundary2*, 29/2 (2002), 1–28.

traced to some relatively specific misleading philosophical picture of what knowledge, mind, meaning, or reality *must* be like—on pain of some Bad Consequence. The best anti-metaphysical strategy is then to diagnose and dissolve that underlying misconception, thereby relieving the felt pressure that had made a metaphysical response seem possible because necessary. McDowell reads Kant and Hegel as already engaged in enterprises with this diagnostic-therapeutical, anti-metaphysical shape. As far as the general issue is concerned, I think this is an anti-metaphysical attitude, and a template for arguing against metaphysical programs, rather than an argument as such.

7

What most gives me pause about the commitments underlying programs of the sort I am calling 'metaphysical' is that they essentially require us to quantify over all possible vocabularies. Universal base languages are base languages from which *every* vocabulary that is legitimate in some sense (specific to the metaphysical program) can be elaborated as a target vocabulary. I have my doubts about that notion. It is not that I am confident that no sense can be made of the notion of <u>all possible vocabularies</u>. It is rather that I do not think it comes with a clear sense. If it is to make sense, we must give it a sense. And I don't know how to do that. As a graduate student, I attended a seminar offered by the great anthropologist Clifford Geertz. In the first session, in the course of defining 'culture', the topic of anthropology, he defined a *language* as "a symbol system in which everything can be said." In a question I expressed my natural curiosity as to what he meant by that "everything." He declined further comment—wisely, perhaps. I'd still like to know what one might mean by it. The worry is that no definite or determinate totality is being delineated. Maybe new vocabularies become possible all the time. This issue arises equally, and for the same reason for talk of all possible *facts*. Following Frege, I understand facts as true thoughts—in the sense of true thinkables, rather than true thinkings, of course. And I understand what is thinkable to be what is claimable, what is expressible in some vocabulary. So there are as many facts as there are true statements in any vocabulary. (I don't object to people using 'fact' so as to allow for the possibility of facts not expressible in any vocabulary. But they must undertake the labor of making

sense of that more capacious notion, and showing that there is real work for it to do.) So talking about "all the facts" and talking about "all possible vocabularies" involves a common set of commitments, ones about which I am uneasy. (I am equally uncomfortable with quantification over all "objects." objects : sortals :: facts : claims, all depending on what possible vocabularies there are.)

One might think in this connection about the third proposition of the *Tractatus*: "The world is determined by the facts, and by their being all the facts" (1.11). Wittgenstein rejects the idea of facts that cannot be expressed: for any fact, there is some vocabulary in which it can be stated (clearly). To talk about "all the facts" is then to talk about everything expressible in *any* vocabulary. The *Tractatus* is predicated on there being some *one* vocabulary in which everything sayable at all can be said. It is in this sense a work of metaphysics. And the claim I've cited is a metaphysical claim. Nothing but embarrassment results if we ask about the official status of the second conjunct of this proposition in the botanization of the *Tractatus*. That a specified collection of facts is *all* the facts is not itself one of those facts, nor is it a fact of that kind. But facts are what can be said. Wittgenstein officially denies that we can even say that there *are* facts, or how many there are (4.1272). Is it, then, something that is *shown*? But how? The awkwardness here shows how the issue of the intelligibility of quantifying over all vocabularies arises in this particular setting. But cognate difficulties attend claims such as "All the facts are physical facts." Putting aside worries about the boundaries of "physical facts" (it is a very implausible claim if one means *contemporary* physics, and it is hard to say what one means by something like "eventual" or "ideal" physics in a way that remains plausible without circularity) and the sense of "are," the question of how to give a definite sense to "all the facts" in such a way as to make the claim plausible without being question-begging remains. I just don't know what we are saying when we talk this way.

Happily, I do not need to assess the success of an objection along these lines to the viability of metaphysical projects in order to address the objection to the classical project of analysis that stems from noting the metaphysical character of its core programs of empiricism and naturalism. For I reject the other limb of the objection: the conditional claim that *if* the core programs are defective, because metaphysical, *then* the analytic enterprise is revealed as defective or pointless. For one need not endorse any

commitment to *globally* privileging some vocabularies as base vocabularies with respect to *all* vocabularies—which are legitimate in one sense or another only if and insofar as they can be reconstructed as target vocabularies elaborated from those base vocabularies—in order to vindicate the project of *locally* identifying *particular* cases where the base-vocabulary/target-vocabulary relation can be shown to obtain. For there is a distinctive kind of philosophical understanding that consists in practical mastery of the relations among vocabularies that become visible when we trace various possible base→target relations through the (admittedly, in general, ill-defined) field of vocabularies. As an example, functionalism, and its specifically computational AI species, is an attempted *local* privileging of one vocabulary over another: functional vocabulary over intentional vocabulary. It has no pretensions at all to *global* privileging of functional vocabulary, to its being a *universal* base vocabulary. Though I have offered some reasons for skepticism about even the most plausible, pragmatist, version of AI-functionalism about intentionality, I think that a suitably broadly construed functionalism is something like the only straw floating as a prospect for an account, in other terms, of intentionality.[18] Of course, it may well be, as some Wittgensteinians, for instance, think, that intentional discourse can only be understood in its own terms, and not in terms of some other vocabulary. For reasons I discuss above, that does not mean that understanding would totter. Analytic (algorithmic, algebraic) understanding is not the only kind of genuine philosophical understanding, and it is not always available.

David Lewis propounded a view of philosophy that was inspiring to me when I was his student, and inspires me still. He thought that what philosophers should do is lay down a set of premises concerning some topic of interest as clearly as possible, and extract consequences from them as rigorously as possible. Having done that, one should lay down another, perhaps quite different set of premises, and extract consequences from them as rigorously as possible. The point was not in the first instance to endorse the conclusions of any of these chains of reasoning, but to learn our way about in the inferential field they all defined, by tracing many overlapping, intersecting, and diverging paths through the terrain. That is how we would learn what difference it would make, in various

---

[18] I say something about this in the first chapter of *Tales of the Mighty Dead*.

contexts, if we were to endorse some claim that figures as a premise in many of the inferences, and what might entitle us to a claim that shows up as a consequence in many of the inferences. Actually plumping for and defending any of these theses is then a subsequent, parasitic, and substantially less important stage of the process. The principal aim is not belief, but understanding.

One thing that was liberating and exhilarating about this metaphilosophical attitude is that Lewis accordingly didn't care much what reasons one had for starting with one set of premises rather than another. He was entirely open to, and indeed eager to, turn his awe-inspiring intellect to following out the consequences of even the wackiest of claims. Pavel Tichy visited while I was still in graduate school at Princeton, and he was then gripped by just such an obsession. He had somehow gotten the idea that there was some finite number n such that it was a necessary truth that there were exactly n things in each possible world, and that there were no further constraints on transworld identification of objects or counterparts. I don't know how he thought he could count objects (Lewis later had what is at least a responsive answer to this question from a physicalist point of view). And I can't imagine what would make one think his axiom true. But Lewis was perfectly willing to figure out the sort of modal logic and metaphysics one would get on that assumption. You never know where you might learn something.

I think this is the spirit in which we should think about semantic relations between different vocabularies. It is worth seeing how and to what extent different target vocabularies can be elaborated from various base vocabularies—including, and perhaps especially, with pragmatic detours through the specifications of practices-or-abilities necessary or sufficient to deploy those vocabularies—because that is a way of coming practically to know our way around those vocabularies, our discursive practices, and the subject matters they make it possible for us to talk and think about. Exploring all the available paths between landmarks is a way, perhaps the only way, of learning to find our way around these woods, acquiring a practical conceptual mastery of the many aspects of discursive practice, and their relations to one another. Let me repeat that I am not claiming this is all there is to philosophy, or that this is the only way to do philosophy. But this sort of semantic analysis yields one valuable kind of understanding that is potentially of great philosophical value.

Notice that, on this view, one's exploration of semantic relations (including pragmatically mediated ones) among vocabularies of antecedent philosophical interest need not be motivated by some global, monolithic program, such as empiricism or naturalism. The merit or benefit of the analytic project is not hostage to such programs, for the distinctive kind of understanding it aims at is not. That understanding can be well served by accumulating particular, local connections that support no antecedent global program and perhaps could be predicted by none (both of which are true of most of the applications of meaning-use analysis I consider in these lectures). Nor must the search for such semantic relations among vocabularies and the discursive practices-or-abilities they specify or that deploy them be motivated by some deep-seated philosophical anxiety or puzzlement, the proper deflating diagnosis of which then exhibits or renders the task of exploring those relations otiose. Simple curiosity, the desire to deepen our understanding, can suffice as much for this sort of philosophical theorizing as for the empirical scientific variety. Indeed, as Kuhn has taught us in the latter case, it really does not matter *why* the scientists do what they do, since the institution can ensure that so long as they act professionally, the result will be to extend our knowledge and deepen our understanding. So we might strive to make it be in philosophy—a light and harmless sort of motivational scientism.

8

With this thought on the table, I want to return briefly to the issue of metaphysics. Taking our cue from Geertz, we might think of metaphysics as the enterprise of crafting a vocabulary in which everything can be said. Now, he might be right that, in the sense I am after, natural languages are autonomous vocabularies in which everything can be said. But "craft" is doing some work in this definition. The metaphysician aims to *construct* a *technical, artificial* vocabulary with that same expressive power. Why? The greater control that regimentation gives vocabularies whose basic semantics is stipulated—in some other vocabulary, perhaps a natural language (no escaping the need for hermeneutic understanding)—and the rest of whose semantics is computed algorithmically. For we have a distinctively clear sort of understanding of whatever other vocabularies can be elaborated

as target vocabularies from a base vocabulary constructed so as to exhibit this structure. We can get around my earlier worries about the concept of <u>all possible vocabularies</u> by understanding the "everything" regulatively, rather than constitutively. That is, for every vocabulary anyone comes up with, the metaphysician is committed to the favored base vocabulary being adequate, when suitably elaborated, to express what it expresses. We start by trying to codify the vocabularies we have, but acknowledge the commitment to address any more that may come along.

I think metaphysics in this sense is a perfectly reasonable undertaking, and that we potentially have a lot to learn from pursuing it. It is, perhaps, somewhat quixotic—but that is a practical, not a theoretical drawback. If we are to reap the rewards in understanding that engaging in this kind of metaphysical enterprise promises, however, I think it is crucial that it be pursued in the open-minded, pluralistic spirit of Lewis, and not in a more small-minded and exclusionary one. The distinction arises when the metaphysician fails to reconstruct in the favored terms *all* the antecedent uses of *all* the vocabularies it aspires to codify. I take it that such partial failures are inevitable. The task is just too hard, both for practical reasons and for principled ones. It is probably too much to ask even that for every target vocabulary one find some other regimented base vocabulary from which it can be elaborated. Finding some one regimented base vocabulary in terms of which every such target vocabulary can be reconstructed is far more difficult and unlikely. That is why it is to some extent a quixotic quest. But for the same sort of reasons that led to Popper's methodological recommendation to endorse the strongest, most easily falsifiable theory not already falsified by the evidence, it can make sense to pursue the quantificationally more difficult goal. This is because, to the limited extent that one does succeed, one finds out more both about the metaphysical base vocabulary and about the target vocabularies to which it turns out either to be expressively adequate, or not to be expressively adequate. Traditional metaphysics treats the distinction as invidious. It denigrates and dismisses what resists formulation in its favored terms as ontologically second class: as unreal, as mere appearance. So, for Leibniz, relations and evil are unreal, relegated to the phenomenal realm of appearance. Later metaphysicians found themselves similarly rejecting as unreal such phenomena as time. (One of my favorite mind-benders is the attempt of some British idealists to dismiss finitude as an illusion, the effect produced by the infinite

Absolute — what there really is — on poor incapable … merely finite minds.) In a more contemporary semantic key, the term of disapprobation may be "unintelligible" rather than the ontological "unreal." That, at any rate, is the way the logical empiricists talk, and admirers of theirs such as Quine continued the practice (modal logic merely "engenders an illusion of understanding" of modal vocabulary, and intentional vocabulary is merely apparently coherent). (Nineteenth-century idealism, which more or less equates the unreal with the unintelligible, may be thought of in this respect as a transitional phase.)

But such a mean-spirited, suspicious, begrudging, exclusionary attitude is not the only one possible. One might instead take the great positive payoff of a particular metaphysical effort to consist in no small part in the particular line it draws between what target vocabularies (and which bits of each) can, and which cannot, be captured with the expressive resources of the base vocabulary whose metaphysical credentials are being explored. Each regimented base vocabulary, we might suppose, will determine a different boundary between the (relatively) expressible and inexpressible. Suppose we found out (it wouldn't surprise me) that there is simply no way to say in the language of fundamental physics, no matter how it is elaborated with the resources of the most advanced logic, what Samnel Beckett said when he said, "I can't go on. … I'll go on," or what Richard Nixon said when he said, "I am not a crook," or what the Buddha said when he told the hot-dog man, "Make me one with everything." We don't need to say that they didn't say anything, or to pretend that we can't understand what they did say, we don't need to deny that there is such a thing as going on, being a crook, and so on, to learn something about saying things from the relation between their vocabularies and that of fundamental physics that shows up in this expressive mismatch. Indeed, I think we learn more if we do not go on to adopt the wholly optional dismissive attitude. Our slogan should be "Metaphysical discrimination without denigration." And just as Lewis thought it essential that we draw consequences rigorously from *many* sets of premises, so as to learn our way around by taking many different paths through the terrain, so the virtues of the metaphysical enterprise will manifest themselves most fully if we try out *many* different possible metaphysical base vocabularies. (A side benefit of adopting this plan is that we then need not resign ourselves to living out our lives oppressed by the steady drip, drip, drip of naturalistic semantics and ontology.)

My characterization of metaphysics transposes what is normally thought of in *ontological* terms into a *semantic* key. These versions can be thought of as related to one another on the model of Carnap's *material* and *formal* modes. I've described the metaphysical project in metalinguistic terms. It might seem that the translation back and forth between these two ways of talking is so straightforward that it is perverse to flout ordinary philosophical usage by insisting on the metalinguistic version here. But in this case there is a significant asymmetry between them. Indeed, I think the asymmetry here reveals something important about the Carnapian dyad that we might otherwise not have seen: a new justification, from the side of pragmatism, for the characteristically analytic preference for the formal mode. For thinking of the metaphysical enterprise in semantic terms, as seeking to establish distinctive sorts of relations among vocabularies, opens up the possibility of considering in this case, too, *pragmatically mediated* semantic relations between vocabularies, in addition to the traditional kind. In particular, we can lay alongside the aspiration to find a vocabulary in which everything can be said, the aspiration to find one in which one can say everything one must be able to *do* in order to say anything, that is to *use* any vocabulary whatsoever. This is just the idea of a <u>universal *pragmatic* metavocabulary</u>.

The sort of illumination one would get from succeeding at the task of constructing a regimented *de facto* universal pragmatic metavocabulary is not exactly the same as that which one would get from succeeding at the task of constructing a regimented vocabulary whose expressive power encompassed that of all the vocabularies we could come up with to try it out on. But they would evidently be complementary forms of understanding: one telling us what we can *say*, and the other what we must *do* to say it. Further, any adequate pragmatic metavocabulary for a semantically adequate metaphysical vocabulary would *be* a universal pragmatic metavocabulary. I have already suggested, however, that the real payoff from the metaphysical enterprise should not be thought of as consequent upon the anticipation of complete success at producing a regimented semantically expressively universal vocabulary. In place of such a wholesale cognitive reward, we should think of the accumulation of retail rewards. Each only partially successful try at a universal metaphysical vocabulary draws a line between those antecedent vocabularies it can reconstruct, and those it cannot. And each such endeavor will draw a *different* line. The lesson I drew

from the young David Lewis's methodologically principled polymorphous theoretical promiscuity is that a valuable kind of understanding consists in the sort of knowing our way about secured only by multiplying the criss-crossing of concrete ways of drawing the boundary between the expressible and the inexpressible, not globally and absolutely, but locally and relative to specific base vocabularies. The same will hold true of attempts to construct regimented universal pragmatic metavocabularies: their value lies in the details of their only partial successes, in where, specifically, they fail, and in how the line between partial success and partial failure varies as we try out quite different candidate base vocabularies. Here one thinks of the parable with which Hempel closes "Empiricist Criteria of Cognitive Significance," in which a dying father tells his sons a vast treasure is buried in their vineyard. Only many years later do they realize that their energetic but fruitless digging in search of the imaginary trove has led to their turning over the soil in just the way needed to ensure that their vines flourish. The romantic dream of total transformation by a single, magical find inspires the hard, unexciting daily work that gradually lays up the real treasure.

The parallel between the metaphysical goal of crafting a regimented universally expressive vocabulary and that of constructing a universal pragmatic metavocabulary, the genus of which these two tasks are species, is invisible if we think of metaphysics exclusively in ontological terms. The wider perspective is available only if we construe it semantically, in terms of relations between vocabularies. So viewed, they show up as complementary, corresponding to elements not only of the dimension defined by the semantic/pragmatic distinction, but also of the object-ive/subjective dimension: what is talked about and talking about it. They address objective-ontological and subjective-practical sides of the coin of discursiveness. Although both these sorts of totalizing project come into view from the vantage point of meaning-use analysis, what I am doing in the body of these lectures is neither metaphysics nor an attempt to construct a universal pragmatic metavocabulary. Rather, I am looking for a different kind of metavocabulary. It is at a higher level, making it possible to express crucial structural features of the *relations* between the dimensions of discursiveness they take as their targets, relations between what is said and what is done.

9

I have devoted this Afterword to responding to the suggestion that it detracts from my enterprise to frame it as a way of continuing and extending the classical twentieth-century project of philosophical analysis, because that project is a degenerating research program whose basic orienting commitments no longer deserve our allegiance. Although, as I acknowledged at the outset, that framing is not an essential element of the meaning-use analysis which it is my primary purpose to introduce here, there is another sense in which associating it with the analytic tradition is not merely optional. For, even apart from that way of motivating it, features that are intrinsic to my project place it squarely within the analytic tradition. (Sometimes you get to choose your tradition, and sometimes it chooses you.) I am, after all, principally concerned with developing a regimented vocabulary for expressing and discussing semantic relations among vocabularies, including pragmatically mediated ones. That is an analytic project, at least relative to one way of distilling an essence out of that multifarious tradition, one way of retrospectively rationally reconstructing it so as to make or find a common project that then becomes visible as having been implicit in it all along. My own philosophical interests focus on discursiveness, and in particular its semantic dimension. So I tend to view the history of philosophy—not just of analytic philosophy, but also Kant and Hegel and their early modern ancestors—through semantic spectacles, whose lenses throw into highest relief philosophers' accounts of the nature of conceptual content and its place in discursive cognitive and practical activity. I think there is much to recommend this perspective on the philosophical tradition.[19] But I am aware that it may appear to many as just another instance of the methodological carpenter's rule: to the man who only has a hammer, the whole world looks like a nail. (More gender-neutral versions of this slogan seem somehow to miss something central about it.) Be that as it may, in a straightforward sense, to construe philosophy's task in semantic terms (including, as I have just done, the traditionally ontological task of metaphysics) is to adopt a characteristically analytic perspective.

---

[19] Which, of course, by no means begins with *Tales of the Mighty Dead*. Alberto Coffa's thought-provoking *The Semantic Tradition from Kant to Carnap* (Cambridge University Press, 1993) is an early index text helping to define this historiographical perspective.

The project pursued here also places itself squarely in the analytic tradition in the centrality and importance it accords to *logical* vocabulary. Trying to get clear about the expressive role distinctive of logical vocabulary is one of the cardinal applications of the general metaconceptual apparatus deployed here. It provides the template on which the analyses of alethic modal and deontic normative vocabulary are modeled. And one of the principal insights gleaned from thinking about pragmatically mediated semantic relations is the new vindication of the semantic logicism characteristic of the classical project of philosophical analysis.

Further, the core idea I am pursuing in these lectures is itself a new kind of analysis: meaning-use analysis. It is generically like the original, narrower, logicist notion, in that it permits the exact, algebraic specification of relations among vocabularies, including those that have traditionally been of interest to philosophers. Complex meaning-use relations are built up from the basic ones combinatorially and recursively, in a way vividly and perspicuously represented by meaning-use diagrams. So it can be settled exactly what one is committed to by, and what one needs to show in order to entitle oneself to, the assertion that various vocabularies and associated sets of discursive practices-or-abilities stand in some specific complex meaning-use relation. Such a regimented, algebraic meta(meta)vocabulary evidently is a generalization of what classical philosophical analysis aimed at, widening the purely semantic scope of the tradition to incorporate relations between the meanings expressed by various vocabularies and the practices-or-abilities of deploying and specifying them.

Finally, among the senses in which a work can belong to the tradition of analytic philosophy is a stylistic one. Analytic philosophers do things such as distinguish six sorts of objections one might have to the analytic project, five senses of 'scientism', four ways in which a project belongs to the analytic tradition, two sorts of conclusion one could draw from the only partial success of a metaphysical project, and so on. In this sense, too, the current project reveals itself throughout as belonging to the analytic tradition. But this is a style of philosophical writing and thinking that long antedated twentieth-century philosophical analysis, for it is fully on display already in Aristotle, in Kant, in Husserl, and in the Heidegger of *Sein und Zeit* and the *Grundprobleme*.

10

I want to close by briefly addressing another *ad hominem* question I have found is often raised by this material: what is the relation between meaning-use analysis and the view developed in *Making It Explicit*? The short answer is that these are distinct projects. Meaning-use analysis does not depend at all on any of the principle theses of *Making It Explicit*: not its normative pragmatics, not its inferentialist semantics, not its account, in terms of those, of what is expressed by representational locutions. That having been said, there are two sorts of connection. First, the theoretical apparatus I present here did grow out of my thinking about what I was doing in *Making It Explicit*—in particular, about how semantic considerations were entangled, often fruitfully, with pragmatic ones. Again and again I found myself needing sharper analytic tools in order to make clear the space of theoretical and explanatory options, and the considerations that told in favor of one or another of them. Meaning-use analysis strives to be neutral as far as the more detailed accounts of meaning, use, and the relations between them, to which it is applied. It aims to be adequate to express Dummett's views, Davidson's, David Lewis's, or Stalnaker's—those, indeed, of any of the relatively small number of philosophers who are explicitly concerned with what it is thinkers and talkers must *do* (the practices they must engage in or the abilities they must exercise) in order thereby to count as associating semantically relevant whatsises (whatever semantic interpretants the semantic portion of the theory assigns) with their locutions and performances. But its home proving-ground was the account of discursive practice and conceptual content presented in *Making It Explicit*.

Second, the examples to which meaning-use analysis is applied in these lectures, and so the specific lessons about philosophically interesting vocabularies it is used to extract, are often ideas that are rooted in or familiar from my earlier work. The discussion in Lecture 2 of the expressive role characteristic of logical vocabulary, for instance, will be immediately recognizable to readers of *Making It Explicit* and *Articulating Reasons*.[20] The apparatus of meaning-use analysis makes it possible to develop and

---

[20] Robert Brandom, *Articulating Reasons: An Introduction to Inferentialism* (Harvard University Press, 2000).

clarify with hitherto impossible precision the inferentialist-expressivist line of thought about logic presented there. Again, the two-sorted deontic normative pragmatics, in terms of <u>commitment</u> and <u>entitlement</u>, and the three-sorted botanization of consequential-inferential semantic relations in terms of which the semantically primitive notion of <u>incompatibility</u> is introduced in Lecture 5 are just those motivated and discussed in *Making It Explicit* and *Articulating Reasons*. The subsequent exposition of incompatibility semantics for alethic modality redeems a promissory note issued in those works. And further examples are not far to seek.

This is not at all because the metatheoretic machinery of these lectures makes sense only in the context of those more particular ground-level views. My conception of the rhetorical shape I wanted the lectures to take is that I would move back and forth between introducing more of the general concepts of meaning-use analysis and making them more definite and showing their importance by applying them to particular cases—to issues concerning vocabularies of antecedent philosophical interest. The more suggestive and illuminating the results of such applications turn out to be, the more reason there is to think that the meta-metavocabulary of meaning-use analysis is cutting at important joints, making visible structures of general significance. The value of the result was supposed to be something like the vector product of what we learned from the applications and what we learn about the apparatus by applying it. Even those with no antecedent interest in relations between meaning and use might find that they learned a lot about issues they did care about (logic, AI, modality, intentionality …) by broadening their horizons to consider pragmatically mediated semantic relations among vocabularies. In looking for such payoffs, I naturally began with topics I already had something to say about, described in a vocabulary articulating a conceptual scheme I had already developed and deployed to some good effect elsewhere, where it seemed that additional analytic power might yield substantial further insights. The real test, of course, will be the extent to which others can usefully apply the ideas introduced here to further problems, vocabularies, and discursive phenomena, described in other terms, to produce promising and fruitful philosophical perspectives, approaches, ideas, and understandings.

Bibliography

Anscombe, G. E. M. (1975). "The First Person," in Samuel Guttenplan (ed.), *Mind and Language*. Oxford: Clarendon Press.

Bouveresse, Jacques (2000). "Reading Rorty: Pragmatism and its Consequences," in Robert Brandom (ed.), *Rorty and His Critics*. Malden, Mass.: Blackwell.

Brandom, Robert (1983). "Asserting," *Noûs*, 17/4 (November), 637–50.

—— (1994). *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, Mass.: Harvard University Press.

—— (2000). *Articulating Reasons: An Introduction to Inferentialism*. Cambridge, Mass.: Harvard University Press.

—— (2001). "Modality, Normativity, and Intentionality," *Philosophy and Phenomenological Research*, 63/3 (November), 587–609, with a comment by Gideon Rosen.

—— (2002). *Tales of the Mighty Dead: Historical Essays in the Metaphysics of Intentionality*. Cambridge, Mass.: Harvard University Press.

—— (2002). "Pragmatics and Pragmatisms," in James Conant and Urszula M. Zeglen (eds.), *Hilary Putnam: Pragmatism and Realism*. New York: Routledge. Translated as "Pragmatik und Pragmatismus," in M. Sandbothe (ed.), *Die Renaissance des Pragmatismus*. Weilerswist: Velbrück Wissenschaft, 2000.

—— (2002). "When Philosophy Paints Its Blue on Gray: Irony and the Pragmatist Enlightenment," *boundary2*, 29/2, 1–28.

—— (2004). "Hermeneutic Practice and Theories of Meaning," *SATS: Nordic Journal of Philosophy*, 5/1, 5–26.

Castañeda, H. N. (ed.) (1975). *Action, Knowledge, and Reality*. Indianapolis: Bobbs-Merrill.

Coffa, Alberto (1993). *The Semantic Tradition from Kant to Carnap*. New York: Cambridge University Press.

Dewey, John (1972 [1896]). "The Reflex-Arc Concept in Psychology," in Jo Ann Boydston (ed.), *John Dewey: The Early Works 1882–1898*, vol. 5. Carbondale: Southern Illinois University Press.

—— (1981 [1925]) *Experience and Nature*, in Jo Ann Boydston (ed.), *John Dewey: The Later Works 1925–1953*, vol. 1. Carbondale: Southern Illinois University Press.

Dreyfus, Hubert L. (1997). *What Computers Still Can't Do*. Cambridge, Mass.: MIT Press.

Fine, Arthur (1996). "The Natural Ontological Attitude," in D. Papineau (ed.), *The Philosophy of Science*. Oxford: Oxford University Press.

Fodor, Jerry A. (1987). "Modules, Frames, Fridgeons, Sleeping Dogs and the Music of the Spheres," in Z. Pylyshyn (ed.), *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*. Norwood, NJ: Ablex.

—— and Lepore, Ernest (2002). *The Compositionality Papers*. New York: Oxford University Press.

Geach, Peter, T. (1969). *God and the Soul*. London: Routledge.

Grim, Patrick, Mar, Gary, and Williams, Peter (eds.) (2003). *The Philosopher's Annual*, vol. 26. Atascadero, Calif.: Ridgeview Publishing Co.

Haugeland, John (1989). *Artificial Intelligence: The Very Idea*. Cambridge, Mass.: MIT Press.

Hegel, Georg W. F. (1979 [1807]). *Phenomenology of Spirit*, A. V. Miller (trans.). Oxford: Oxford University Press.

—— (1990 [1812–16]). *Science of Logic*, A. V. Miller (trans.). New York: Humanity Books.

Hempel, Carl G. (1970 [1950]). "Empiricist Criteria of Cognitive Significance," in *Aspects of Scientific Explanation*. New York: Free Press.

—— (1970 [1958]). "The Theoretician's Dilemma," in *Aspects of Scientific Explanation*. New York: Free Press.

Kneale, William and Kneale, Martha (1962). *The Development of Logic*. Oxford: Oxford University Press.

Kripke, Saul (1972). "Naming and Necessity," in Donald Davidson and Gilbert Harman (eds.), *Semantics of Natural Language*. Dordrecht: Reidel.

Kuklick, Bruce (2004). "Philosophy at Yale in the Century after Darwin," *History of Philosophy Quarterly*, 21, 313–36.

Leibniz, Gottfried (1929 [1685]). "Machina arithmetica in qua non additio tantum et subtractio sed et multiplicatio nullo, divisio vero paene nullo animi labore peragantur," Mark Kormes (trans.), in D. E. Smith (ed.), *A Source Book in Mathematics*. New York: McGraw-Hill.

Lewis, David (1972). "General Semantics," *Synthese*, 22, 18–67.

—— (1979). "Attitudes *De Dicto* and *De Se*," *The Philosophical Review*, 88, 513–43.

Macbeth, Danielle (2005). *Frege's Logic*. Cambridge, Mass.: Harvard University Press.

McDowell, John (2001). "*De Re* Senses," in *Meaning, Knowledge, and Reality*. Cambridge, Mass.: Harvard University Press.

MacFarlane, John (2000). "What Does It Mean to Say that Logic is Formal?" PhD dissertation, University of Pittsburgh.

Perry, John (1979). "The Problem of the Essential Indexical," *Noûs*, 13/1 (March), 3–21.

Price, Huw (2003). "Truth as Convenient Friction," *Journal of Philosophy*, 100, 167–90.

—— (2004). "Naturalism without Representationalism," in Mario de Caro and David Macarthur (eds.), *Naturalism in Question*. Cambridge, Mass.: Harvard University Press.

Quine, W. V. O. (2006 [1953]). "Two Dogmas of Empiricism," in *A Logical Point of View: Nine Logico-Philosophical Essays*, 2nd rev. edn. Cambridge, Mass.: Harvard University Press.

Rorty, Richard (1999). "Pragmatism as Anti-authoritarianism," *Revue Internationale de Philosophie*, 53/1, 7–20.

—— (2000). "Universality and Truth," in Robert Brandom (ed.), *Rorty and His Critics*. Malden, Mass.: Blackwell.

Ryle, Gilbert (1950). "'If', 'So', and 'Because'," in Max Black (ed.), *Philosophical Analysis*. Englewood Cliffs, NJ: Prentice Hall.

Schneider, Hans Julius (1992). *Phantasie und Kalkül: Über die Polarität von Handlung und Struktur in der Sprache*. Frankfurt a. M.: Suhrkamp.

Searle, John R. (1997 [1980]). "Minds, Brains, and Programs," in John Haugeland (ed.), *Mind Design II*. Cambridge, Mass.: MIT Press.

Sellars, Wilfrid (1980 [1948]). "Concepts as Involving Laws, and Inconceivable Without Them," in J. Sicha (ed.), *Pure Pragmatics and Possible Worlds: The Early Essays of Wilfrid Sellars*. Reseda, CA: Ridgeview Publishing Co.

—— (1980 [1949]). "Language, Rules, and Behavior," in J. Sicha (ed.), *Pure Pragmatics and Possible Worlds: The Early Essays of Wilfrid Sellars*. Reseda, CA: Ridgeview Publishing Co.

—— (2007 [1953]). "Inference and Meaning," in Kevin Scharp and Robert Brandom (eds.), *In the Space of Reasons*. Cambridge, Mass.: Harvard University Press.

—— (1953). "Is There a Synthetic *A Priori*?" *Philosophy of Science*, 20, 121–38.

—— (1997 [1956]). *Empiricism and the Philosophy of Mind*, edited and with a Study Guide by Robert Brandom. Cambridge, Mass.: Harvard University Press.

—— (1957). "Counterfactuals, Dispositions, and the Causal Modalities," in H. Feigl, M. Scriven, and G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science*, vol. 2. Minneapolis: University of Minnesota Press.

—— (1963). "Phenomenalism," in *Science, Perception, and Reality*. New York: Routledge & Kegan Paul. Reprinted in Kevin Scharp and Robert Brandom (eds.), *In the Space of Reasons*. Cambridge, Mass.: Harvard University Press, 2007.

Sudkamp, Thomas A. (1998). *Languages and Machines*, 2nd edn. Reading, Mass.: Addison Wesley Longman.

Thompson, Michael (forthcoming). *Life and Action*. Cambridge, Mass.: Harvard University Press.

Wilson, Mark (2006). *Wandering Significance*. Cambridge, Mass.: Harvard University Press.

Wittgenstein, Ludwig (2001 [1921]). *Tractatus Logico-Philosophicus*, 2nd edn. New York: Routledge.

—— (1999 [1953]). *Philosophical Investigations*, G. E. M. Anscombe (trans.). Malden, Mass.: Blackwell.

This page intentionally left blank

Subject Index

Name Index