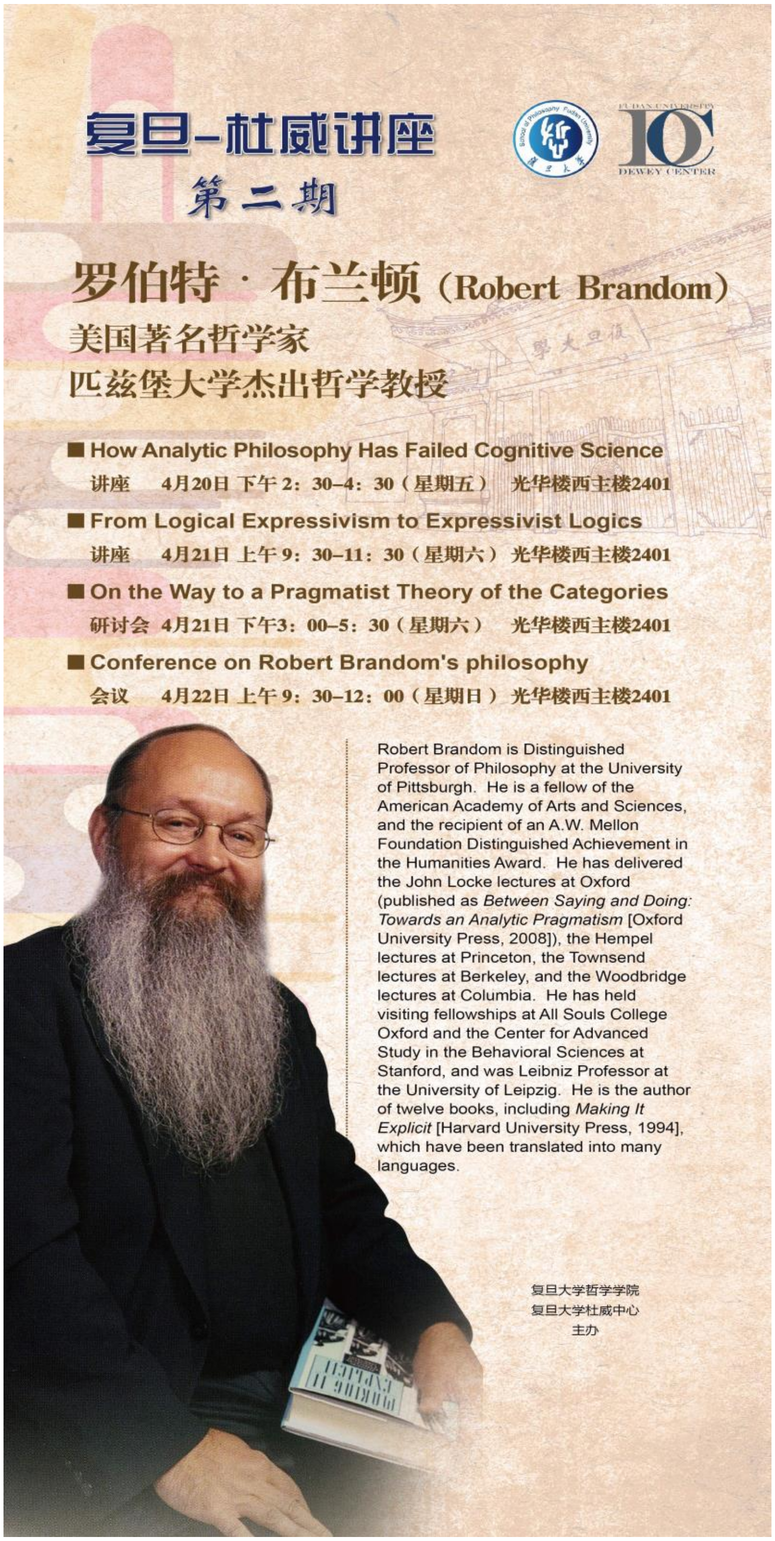
Inferentialism, Normative Pragmatism, and Metalinguistic Expressivism

Robert Brandom’s 2018 John Dewey Lectures at the Dewey Center of Fudan University



复旦-杜威讲座

第二期

罗伯特 · 布兰顿（Robert Brandom）

美国著名哲学家

匹兹堡大学杰出哲学教授

- **How Analytic Philosophy Has Failed Cognitive Science**
  讲座　4月20日 下午 2：30–4：30（星期五）　光华楼西主楼2401
- **From Logical Expressivism to Expressivist Logics**
  讲座　4月21日 上午 9：30–11：30（星期六）光华楼西主楼2401
- **On the Way to a Pragmatist Theory of the Categories**
  研讨会 4月21日 下午3：00–5：30（星期六）　光华楼西主楼2401
- **Conference on Robert Brandom's philosophy**
  会议　4月22日 上午 9：30–12：00（星期日）光华楼西主楼2401

Robert Brandom is Distinguished Professor of Philosophy at the University of Pittsburgh. He is a fellow of the American Academy of Arts and Sciences, and the recipient of an A.W. Mellon Foundation Distinguished Achievement in the Humanities Award. He has delivered the John Locke lectures at Oxford (published as *Between Saying and Doing: Towards an Analytic Pragmatism* [Oxford University Press, 2008]), the Hempel lectures at Princeton, the Townsend lectures at Berkeley, and the Woodbridge lectures at Columbia. He has held visiting fellowships at All Souls College Oxford and the Center for Advanced Study in the Behavioral Sciences at Stanford, and was Leibniz Professor at the University of Leipzig. He is the author of twelve books, including *Making It Explicit* [Harvard University Press, 1994], which have been translated into many languages.

复旦大学哲学学院
复旦大学杜威中心
主办

Table of Contents

Preface

1. How Analytic Philosophy Has Failed Cognitive Science

2. Some Strands of Wittgenstein's Normative Pragmatism,
   and Some Strains of his Semantic Nihilism

3. On the Way to a Pragmatist Theory of the Categories

4. From Logical Expressivism to Expressivist Logics

Inferentialism, Normative Pragmatism, and Metalinguistic Expressivism

Preface:

I think of us as most deeply and essentially *discursive* creatures—that is, concept-using, linguistic beings. The overarching twin goods for us are accordingly practicing the sorts of sapient understanding and unfettered conversation that only language makes possible. On this way of thinking, at the core of philosophy is philosophy of language. Three pillars of my philosophical thought about discursive practice are semantic inferentialism, normative pragmatism, and metalinguistic expressivism. My Dewey lectures are designed to introduce these ideas.

Semantics is the study of linguistic *meaning* and conceptual content. The modern Western philosophical tradition has taken *representation* to be the key concept of semantics. To understand the sort of contentfulness characteristic of sapience, that tradition counsels us to focus on the relation between pictures and what they picture, between signs and what they are signs for. The master-idea of semantic inferentialism is to look instead to *inference*, rather than representation, as the basic concept of semantics. What makes something meaningful or contentful in the sense that matters for sapience (rather than the mere sentience we share with many nonlinguistic animals) is the role that it plays in *reasoning*. The primary vehicle of meaning in this sense is declarative sentences. Those are symbols that can be used to *assert*, *state*, or *claim that* things are thus-and-so. The kind of content they express, “propositional” content, in the philosopher’s jargon, is what can both serve as and stand in need of reasons—that is what can play the role both of premise and of conclusion in *inferences*. Chapter One, the first of these lectures, explains some of the ways in which approaching meaning and content from an inferentialist perspective leads us to ask different questions, and to highlight different features of our cognitive abilities, than the more traditional representationalist perspective does. It tries to give some idea about how this different perspective might be of use, not just in philosophy, but in cognitive science more generally—for instance, in our understanding of how children develop conceptually, our understanding of the nature of the progression from nonlinguistic primates to sapient humans, and our understanding of the challenges and prospects of artificial sapience in non-sentient (because nonliving) computers.

Pragmatics is the study of the *use* of language, the study of discursive *practice*. Where semantics studies what one is *saying* in making a claim, pragmatics studies what one is *doing* in making a claim. Pragmatism in general is the claim that pragmatics is methodologically, conceptually, and explanatorily prior to semantics—that one should understand the meaning or content expressed by linguistic locutions in terms of their use. The later Wittgenstein, who counseled "Don't look to the meaning, look to the use," is a pragmatist in this sense (though he didn't use that term). *Normative* pragmatism is the idea that discursive practice is implicitly, but essentially, and not just accidentally, a kind of normative practice. Discursive creatures live, and move, and have their being in a normative space. What one is doing in making a claim, performing the most fundamental kind of speech act, is *committing* oneself, exercising one's authority to make oneself *responsible*. The commitments one undertakes in claiming (the beliefs one expresses in sincerely asserting something) are ones whose entitlement is always potentially at issue. Understanding someone's utterance is knowing what they have committed themselves to by producing that performance, by saying what they said—as well as knowing what would entitle them to that commitment, and what is incompatible with it. Those commitments, entitlements, and incompatibilities are *inferentially* connected to one another. The space discursive creatures move about in by talking is a space of *reasons*, articulating what would be a reason for or against what. That is what connects normative pragmatism to semantic inferentialism. Chapter Two addresses Wittgenstein's influential normative pragmatism. It aims to distinguish his most important (indeed, revolutionary) insights from a mistaken conclusion that he draws from those insights. For he thought that focusing to begin with on the use of language rather than the meanings expressions acquire by being used in the ways they are—being subject to the sorts of normative appraisals of correctness they are subject to—means that a certain kind of systematic philosophical theorizing about language becomes impossible. This is Wittgenstein's famous theoretical quietism. I argue that it is not required by his pragmatism.

Kant had the idea that in addition to empirical descriptive concepts, whose expressive job it is to describe and explain what goes on in the world around us, there are concepts he called "categorial," whose distinctive expressive job it is to make explicit features of the framework of discursive practices that makes it possible to describe or explain anything at all—to make claims about how things empirically are. My philosophical hero and former colleague, Wilfrid Sellars, following clues he found in the work of the logical empiricist Rudolf Carnap, argues that there are indeed such framework-explicating concepts, that they do work quite differently from ordinary empirical descriptive/explanatory concepts, and that they should be understood as *metalinguistic* in a broad sense. That is, they let us talk about semantics and pragmatics—about what we are saying and doing when we say things. Among the concepts that should be understood as playing this distinctive sort of metalinguistic expressive role are many concepts that have most puzzled philosophers: semantic concepts, normative concepts, logical concepts, and probabilistic concepts among them. One is bound to find them puzzling if one assimilates them to ordinary descriptive concepts, and tries to understand them representationally. Semantic facts, normative facts, logical facts, and probabilistic facts are bound to be puzzling. For in making claims using this sort of vocabulary, we are not representing or describing how the world is. We are doing something quite different, and so saying something quite different. Chapter Three explores this idea, and what might be made of it going forward.

For the past 120 years or more, logic has played an axial role in analytic philosophy. Most of the stories told about why this should be so have turned out to be unsustainable, however. Logical expressivism offers a satisfying rationale for the long-held conviction that logic is of central importance to philosophy. According to this view, logical vocabulary plays a quite distinctive expressive role. It does not describe or represent anything, and it does not provide a canon of right reasoning. Rather, it is the discursive organ of semantic self-consciousness. According to semantic inferentialism, meaning is something like role with respect to material (prelogical) relations of inferential consequence and incompatibility (rational inclusion and exclusion). To understand others' utterances, or even our own thoughts, we must have the ability to distinguish in practice what follows from what. That is a kind of know-how. Logical vocabulary such as the conditional (if___then...) let's us *say that* one claimable content follows from another. Negation lets us *say that* two claimable contents are incompatible. Without the expressive power this logical vocabulary affords, we can in practice take or treat two claimables as standing in consequential or incompatibility relations. But that is something we *do*, not something we can at that point *say.* Logic lets us make explicit the implicit inferential relations in virtue of which nonlogical concepts have the contents they do. Thinking in this way about the expressive role distinctive of logical vocabulary turns out to suggest that we do logic somewhat differently than we traditionally have. Chapter Four discusses logical Expressivism and the new way of doing logic that can be built on this philosophy of Logic.

All these views can be found in my 1994 book, *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Further development of normative pragmatism can be found in my 2011 book *Perspectives on Pragmatism: Classical, Recent, and Contemporary*. Further development of semantic inferentialism can be found in my 2009 book *Reason in Philosophy: Animating Ideas*, and in my 2008 book *Between Saying and Doing: Towards an Analytic Pragmatism*, which also addresses logical expressivism. Further development of pragmatic metalinguistic expressivism can be found in that book, and in my 2015 book *From Empiricism to Expressivism: Brandom Reads Sellars*.

***

For many years there was a tendency among those who studied American pragmatism to identify that tradition with the founding pragmatists of the Golden Age: Charles Sanders Peirce, William James, and John Dewey. It is a great intellectual and institutional virtue of the Dewey Center at Fudan University under the guidance of Professor Yajun Chen that it has taken a broader view. Dewey is rightly seen as philosophically significant principally as the great promoter of pragmatism of his age. But the tradition of pragmatism he championed did not die with him. He was not the last American pragmatist. In particular, Richard Rorty, the great admirer of James and Dewey, revived and reshaped pragmatism in his own time, and made visible W.V.O. Quine, Wilfrid Sellars, and Donald Davidson as heirs and avatars of pragmatism, even though none of them explicitly adopts the label. Chen follows Rorty in understanding the pragmatist tradition as opening a revelatory window into American philosophy more generally.

Bruce Kuklick and Murray Murphy have urged that we think about C. I. Lewis as setting out to synthesize the views of his teachers William James and Josiah Royce, and that we acknowledge the central intellectual role Lewis goes on to play in American pragmatism. Kuklick urges further that we think about the crucial role that C. I. Lewis's pragmatism plays in shaping the thought of his Harvard students Quine and Nelson Goodman. Cheryl Misak has emphasized that alongside the sequence of the classical triumvirate of Peirce, James, and Dewey, we should acknowledge a second substantial pragmatist filiation that runs from Peirce through C. I. Lewis to Wilfrid Sellars.

It can be a worthwhile enterprise to consider these figures from the perspective of American philosophers alone. From my point of view it is important to keep in mind, even when doing so, the substantial influence of German philosophy on these figures—to begin with, the influence of the great German Idealists Kant and Hegel. Both were central figures for Peirce. James's colleague, friend, and intellectual interlocutor Royce was the principal American Hegelian of his generation. Dewey started off as a Hegelian, and, following hints from Peirce, developed his form of pragmatism as a sort of naturalized, post-Darwinian Hegelian holism. C. I. Lewis was a neo-Kantian, whose influence shaped the teaching of Kant's theoretical philosophy in America for generations afterwards. Sellars was above all a student of Kant, who characterized the effect he hoped his work would have as “moving analytic philosophy from its Humean to its Kantian phase.” (Rorty has described my own aim as “moving analytic philosophy from its incipient Kantian to its eventual Hegelian phase.”) A decisive element in the development of Sellars’s mature thought was a Gestalt-shift in which he came to see his hero Rudolf Carnap as a (repressed) neo-Kantian, transposing that tradition into a metalinguistic key. Rorty thought of himself as a foe of Kantian transcendentalism and a friend of Hegelian historicism. And his understanding of pragmatism was as a philosophical movement broad enough to encompass both the early Heidegger and the later Wittgenstein.

All of these influences have been crucial in orienting my own thought. The two most proximal and manifest legacies are that of Richard Rorty, my teacher, Doktorvater, and friend, and that of my former colleague, Wilfrid Sellars. They are the principal conduits through which the influence of all the other figures I have mentioned are conducted. In this way, I see myself as working at the confluence of the James-Dewey stream of American philosophy that Rorty extended and developed and the Peirce-Lewis stream of American philosophy that Sellars extended and developed. One manifestation of this synthetic aspiration is the conjunction of normative pragmatism, a development of Rorty’s views (shaped by what he helped me see in Sellars and Quine and Dewey and Wittgenstein) and semantic inferentialism, a development of Sellars’s views. Pragmatic metalinguistic expressivism—the understanding of crucial philosophical concepts as letting us express explicitly in the object-language what could also be formulated in a pragmatic metalanguage that lets us *say* what we are *doing* in *using* various kinds of linguistic expression—is also inspired by the juxtaposition of Sellarsian and Rortyan ideas. Logical expressivism, and the expressivist approach to logic it motivates, is my way of contributing in a Sellarsian spirit to the Peirce-Lewis-Quine tradition of logic and the philosophy of logic—and bringing it into more intimate conversation with the early Frege (and Frege's student Carnap), on the German side.

Inviting me to deliver the Dewey lectures is accordingly a welcome, appropriate, and propitious expression of Professor Chen's vision of the dialogue that the Dewey Center at Fudan University is opening up with American pragmatism and American philosophy more generally. The occasion of the actual lectures was made more resonant and significant for me by the realization that I and my wife Barbara arrived for the lecture series in Shanghai 99 years—almost to the day—after John Dewey and his wife Alice arrived in Shanghai for their momentous (and it turned out, extended) visit. After a century of remarkable development of both, it is fascinating to speculate about what they would have thought about China, and about American philosophy, as they would find them today. Certainly they would be glad to know that the Dewey Center and the Dewey lecture were on-going in Shanghai.

Inferentialism, Normative Pragmatism, and Metalinguistic Expressivism

Chapter One:

How Analytic Philosophy Has Failed Cognitive Science

I. Introduction

We analytic philosophers have signally failed our colleagues in cognitive science. We have done that by not sharing central lessons about the nature of concepts, concept-use, and conceptual content that have been entrusted to our care and feeding for more than a century.

I take it that analytic philosophy began with the birth of the new logic that Gottlob Frege introduced in his seminal 1879 *Begriffsschrift*. The idea, taken up and championed to begin with by Bertrand Russell, was that the fundamental insights and tools Frege made available there, and developed and deployed through the 1890s, could be applied throughout philosophy to advance our understanding of understanding and of thought in general, by advancing our understanding of concepts—including the particular concepts with which the philosophical tradition had wrestled since its inception. For Frege brought about a revolution not just in *logic*, but in *semantics*. He made possible for the first time a *mathematical* characterization of meaning and conceptual content, and so of the structure of sapience itself. Henceforth it was to be the business of the new movement of analytic philosophy to explore and amplify those ideas, to exploit and apply them wherever they could do the most good. Those ideas are the cultural birthright, heritage, and responsibility of analytic philosophers. But we have not done right by them. For we have failed to communicate some of the most basic of those ideas, failed to explain their significance, failed to make them available in forms usable by those working in allied disciplines who are also professionally concerned to understand the nature of thought, minds, and reason.

Contemporary cognitive science is a house with many mansions. The provinces I mean particularly to be addressing are cognitive psychology, developmental psychology, animal psychology (especially primatology), and artificial intelligence. (To be sure, this is not all of cognitive science. But the points I will be making in this paper are not of similarly immediate significance for such other subfields as neurophysiology, linguistics, perceptual psychology, learning theory, and the study of the mechanisms of memory.) Cognitive psychology aims at reverse-engineering the human mind: figuring out how we do what we do, what more basic abilities are recruited and deployed (and how) so as to result in the higher cognitive abilities we actually display. Developmental psychology investigates the sequence of stages by which those abilities emerge from more primitive versions as individual humans mature. Animal psychology, as I am construing it, is a sort of combination of cognitive psychology of non-human intelligences and a phylogenetic version of ontogenetic human developmental psychology. By contrast to all these empirical inquiries into actual cognition, artificial intelligence swings free of questions about how any actual organisms do what they do, and asks instead what constellation of abilities of the sort we know how to implement in artifacts might in principle yield sapience.

Each of these disciplines is in its own way concerned with the broadly empirical question of how the trick of cognition is or might be done. Philosophers are concerned with the normative question of what counts as doing it—with what understanding, particularly discursive, conceptual understanding consists in, rather than how creatures with a particular contingent constitution, history, and armamentarium of basic abilities come to exhibit it. I think Frege taught us three fundamental lessons about the structure of concepts, and hence about all possible abilities that deserve to count as concept-using abilities.[1] The conclusion we should draw from his discoveries is that concept-use is intrinsically stratified. It exhibits at least four basic layers, with each capacity to deploy concepts in a more sophisticated sense of 'concept' structurally presupposing the capacities to use concepts in all of the more primitive senses. The three lessons that generate the structural hierarchy oblige us to distinguish between:

- concepts that only *label* and concepts that *describe,*
- *ingredient* and *free-standing* conceptual contents, making explicit the distinction between the *content* of concepts and the *force* of applying them, and
- concepts expressible already by *simple* predicates and concepts expressible only by *complex* predicates.

AI researchers and cognitive, developmental, and animal psychologists need to take account of the different grades of conceptual content made visible by these distinctions, both in order to be clear about the topic they are investigating (if they are to tell us how the trick is done, they must be clear about exactly which trick it is) and because the empirical and in-principle possibilities are constrained by the way the abilities to deploy concepts in these various senses structurally presuppose the others that appear earlier in the sequence. This is a point they have long appreciated on the side of basic *syntactic* complexity. But the at least equally important—

---

[1] It ought to be uncontroversial that the last two of the three lessons are due to Frege. Whether he is responsible also for the first is more contentious. Further, I think both it and a version of the second can be found already in Kant. (As I argue in my 2006 Woodbridge Lectures, "Animating Ideas of Idealism: A Semantic Sonata in Kant and Hegel," forthcoming in the *Journal of Philosophy*.) But my aims here are not principally hermeneutical or exegetical—those issues don't affect the question of what we philosophers ought to be teaching cognitive scientists—so I will not be concerned to justify these attributions.

and I would argue more conceptually fundamental—hierarchy of *semantic* complexity has been largely ignored.

II. **First Distinction: From Labeling to Describing**

The Early Modern philosophical tradition was built around a *classificatory* theory of consciousness and (hence) of concepts, in part the result of what its scholastic predecessors had made of their central notion of Aristotelian forms. The paradigmatic cognitive act is understood as classifying: taking something particular *as* being of some general kind. Concepts are identified with those general kinds.

This conception was enshrined in the order of logical explanation (originating in Aristotle's *Prior Analytics*) that was common to *everyone* thinking about concepts and consciousness in the period leading up to Kant. At its base is a doctrine of *terms* or *concepts*, particular and general. The next layer, erected on that base, is a doctrine of *judgments*, describing the kinds of classificatory relations that are possible among such terms. For instance, besides classifying Socrates as human, humans can be classified as mortal. Finally, in terms of those metaclassifications grouping judgments into kinds according to the sorts of terms they relate, a doctrine of *consequences* or *syllogisms* is propounded, classifying valid inferences into kinds, depending on which classes of classificatory judgments their premises and conclusions fall under.

It is the master-idea of classification that gives this traditional order of explanation its distinctive shape. That idea defines its base, the relation between its layers, and the theoretical aspiration that animates the whole line of thought: finding suitable ways of classifying terms and judgments (classifiers and classifications) so as to be able to classify inferences as good or bad solely in virtue of the kinds of classifications they involve. The fundamental metaconceptual role it plays in structuring philosophical thought about thought evidently made understanding the concept of classifying itself a particularly urgent philosophical task. Besides asking what differentiates various kinds of classifying, we can ask what they have in common. What is it one must *do* in order thereby to count as *classifying* something as being of some kind?

In the most general sense, one classifies something simply by responding to it differentially. Stimuli are grouped into kinds by the response-kinds they tend to elicit. In this sense, a chunk of iron classifies its environments into kinds by rusting in some of them and not others, increasing or decreasing its temperature, shattering or remaining intact. As is evident from this example, if classifying is just exercising a reliable differential responsive disposition, it is a ubiquitous feature of the inanimate world. For that very reason, classifying in this generic sense is not an attractive candidate for identification with conceptual, cognitive, or conscious activity. It doesn't draw the right line between thinking and all sorts of thoughtless activities. Pan-psychism is too high a price to pay for cognitive naturalism.

That need not mean that taking *differential responsiveness* as the genus of which *conceptual classification* is a species is a bad idea, however. A favorite idea of the classical British empiricists was to require that the classifying response be entering a *sentient* state. The intrinsic characters of these sentient states are supposed to sort them immediately into repeatable kinds. These are called on to function as the *particular* terms in the base level of the neo-Aristotelian logical hierarchy. *General* terms or concepts are then thought of as sentient state-kinds derived from the particular sentient state-kinds by a process of *abstraction*: grouping the base-level sentient state-repeatables into higher-level sentient state-repeatables by some sort of perceived *similarity*. This abstractive grouping by similarity is itself a kind of classification. The result is a path from one sort of consciousness, sentience, to a conception of another sort of consciousness, sapience, or conceptual consciousness.

A standing felt difficulty with this empiricist strategy is the problem of giving a suitably naturalistic account of the notion of sentient awareness on which it relies. Recent information-theoretic accounts of representation (under which heading I include not just Fred Dretske's theory, which actually goes by that name, but others such as Jerry Fodor's asymmetric counterfactual dependence and nomological locking models[2]) develop the same basic differential responsiveness version of the classic classificatory idea in wholly naturalistic modal terms. They focus on the information conveyed about stimuli—the way they are grouped into repeatables—by their reliably eliciting a response of one rather than another repeatable response-kind from some system. In this setting, unpalatable pan-psychism can be avoided not, as with traditional empiricism, by insisting that the responses be sentient states, but for instance by restricting attention to flexible systems, capable in principle of coming to encode many different groupings of stimuli, with a process of *learning* determining what classificatory dispositions each one actually acquires. (The classical American pragmatists' program for a naturalistic empiricism had at its core the idea that the structure common to evolutionary development and individual learning is a Test-Operate-Test-Exit negative feedback process of acquiring practical habits, including discriminative ones.[3])

---

[2] Dretske, Fred: *Knowledge and the Flow of Information* (MIT Press—Bradford, 1981), Fodor, Jerry: *A Theory of Content* (MIT Press—Bradford, 1990).

Classification as the exercise of reliable differential responsive dispositions (however acquired) is not by itself yet a good candidate for *conceptual* classification, in the basic sense in which applying a concept to something is *describing* it. Why not? Suppose one were given a wand, and told that the light on the handle would go on if and only if what the wand was pointed at had the property of being *grivey*. One might then determine empirically that speakers are grivey, but microphones not, doorknobs are but windowshades are not, cats are and dogs are not, and so on. One is then in a position reliably, perhaps even infallibly, to apply the *label* 'grivey'. Is one also in a position to *describe* things *as* grivey? Ought what one is doing to qualify as applying the *concept* grivey to things? Intuitively, the trouble is that one does not know what one has found out when one has found out that something is grivey, does not know what one is taking it to be when one takes it to be grivey, does not know what one is describing it *as*. The label is, we want to say, uninformative.

What more is required? Wilfrid Sellars gives this succinct, and I believe correct, answer:

---

[3] I sketch this program in the opening section of "The Pragmatist Enlightenment (and its Problematic Semantics)" *European Journal of Philosophy*, Vol 12 No 1, April 2004, pp. 1-16.

> It is only because the expressions in terms of which we describe objects, even such basic expressions as words for the perceptible characteristics of molar objects, locate these objects in a space of implications, that they describe at all, rather than merely label.[4]

The reason 'grivey' is merely a *label*, that it classifies without informing, is that nothing *follows* from so classifying an object. If I discover that all the boxes in the attic I am charged with cleaning out have been labeled with red, yellow, or green stickers, all I learn is that those labeled with the same color share *some* property. To learn what they *mean* is to learn, for instance, that the owner put a red label on boxes to be discarded, green on those to be retained, and yellow on those that needed further sorting and decision. Once I know what *follows* from affixing one rather than another label, I can understand them not as *mere* labels, but as *descriptions* of the boxes to which they are applied. Description is classification with *consequences,* either immediately practical ("to be discarded/examined/kept") or for further classifications.

Michael Dummett argues generally that to be understood as conceptually contentful, expressions must have not only *circumstances* of appropriate application, but also appropriate *consequences* of application.[5] That is, one must look not only *upstream*, to the circumstances (inferential and non-inferential) in which it is appropriate to apply the expression, but also *downstream* to the consequences (inferential and non-inferential) of doing so, in order to grasp the content it expresses. One-sided theories of meaning, which seize on one aspect to the exclusion of the other, are bound to be defective, for they omit aspects of the use that are essential to meaning. For instance, expressions can have the same circumstances of application, and different consequences of application. When they do, they will have different descriptive content.

---

[4] Pp. 306-307 (§107) in: Wilfrid Sellars: "Counterfactuals, Dispositions, and Causal Modalities" In *Minnesota Studies in the Philosophy of Science, Volume II: Concepts, Theories, and the Mind-Body Problem,* ed. Herbert Feigl, Michael Scriven, and Grover Maxwell (Minneapolis: University of Minnesota Press, 1958), p.225-308.

[5] I discuss this view of Dummett's (from his *Frege: Philosophy of Language* second edition [Harvard University Press 1993], originally published in 1974), at greater length in Chapter Two of *Making It Explicit* [Harvard University Press, 1994], and Chapter One of *Articulating Reasons* [Harvard University Press, 2000].

1]	I will write a book about Hegel,

and

2]	I foresee that I will write a book about Hegel,

say different things about the world, describe it as being different ways. The first describes my future activity and accomplishment, the second my present aspiration. Yet the circumstances under which it is appropriate or warranted to assert them—the situations to which I ought reliably to respond by endorsing them—are the same (or at least, can be made so by light regimentation of a prediction-expressing use of ‘foresee’). Here, to say that they have different descriptive content can be put by saying that they have different truth conditions. (That they have the same assertibility conditions just shows how assertibility theories of meaning, as one-sided in Dummett’s sense, go wrong.) But that same fact shows up in the different positions they occupy in the “space of implications.” For from the former it follows that I will not be immediately struck by lightning, that I will write some book, and, indeed, that I will write a book about Hegel. None of these is in the same sense a consequence of the second claim.

We might train a parrot reliably to respond differentially to the visible presence of red things by squawking “That’s red.” It would not yet be *describing* things as red, would not be applying the concept red to them, because the noise it makes has no significance for it. It does not know that it follows from something’s being red that it is colored, that it cannot be wholly green, and so on. Ignorant as it is of those inferential consequences, the parrot does not grasp the concept (any more than we express a concept by ‘grivey’). The lesson is that even observational concepts, whose principal circumstances of appropriate application are non-inferential (a matter of reliable dispositions to respond differentially to non-linguistic stimuli) must have inferential consequences in order to make possible description, as opposed to the sort of classification effected by non-conceptual labels.

The rationalist idea that the inferential significance of a state or expression is essential to its *conceptual* contentfulness is one of the central insights of Frege’s 1879 *Begriffsschrift* (“concept writing”)—the founding document of modern logic and semantics—and is appealed to by him in the opening paragraphs to define his topic:

> ...there are two ways in which the content of two judgments may differ; it may, or it may not, be the case that all inferences that can be drawn from the first judgment when combined with certain other ones can always also be drawn from the second when

> combined with the same other judgments…I call that part of the content that is the same in both the conceptual content [begriffliche Inhalt]. [6]

Here, then, is the first lesson that analytic philosophy ought to have taught cognitive science: there is a fundamental meta-conceptual distinction between classification in the sense of *labeling* and classification in the sense of *describing*, and it consists in the *inferential* consequences of the classification: its capacity to serve as a premise in inferences (practical or theoretical) to further conclusions. (Indeed, there are descriptive concepts that are purely *theoretical*—such as gene and quark—in the sense that in addition to their inferential consequences of application, they have *only* inferential *circumstances* of application.) There is probably no point in fighting over the minimal circumstances of application of the concepts concept and conceptual. Those who wish to lower the bar sufficiently are welcome to consider purely classificatory labels as a kind of concept (perhaps so as not to be beastly to the beasts, or disqualify human infants, bits of our brains, or even some relatively complex computer programs wholly from engaging in conceptually articulated activities). But *if* they do so, they must *not* combine those circumstances of application with the consequences of application appropriate to genuinely *descriptive* concepts—those that *do* come with inferential significances downstream from their application.

Notice that this distinction between labeling and describing is untouched by two sorts of elaborations of the notion of labeling that have often been taken to be of great significance in thinking about concepts from the classical classificatory point of view. One does not cross the boundary from labeling to describing just because the reliable capacity to respond differentially is *learned*, and in that sense flexible, rather than *innate*, and in that sense rigid. And one is likewise developing the classical model in an orthogonal direction insofar as one focuses on the metacapacity to learn to distinguish arbitrary Boolean combinations of microfeatures one can already reliably discriminate. From the point of view of the distinction between labeling and describing, that is not yet the capacity to form *concepts*, but only the mastery of *compound* labels. That sort of structural articulation upstream has no *semantic* import at the level of description until and unless it is accorded a corresponding inferential significance downstream.

---

[6] Frege, *Begriffsschrift* (hereafter *BGS*), section 3. The passage continues: "In my formalized language [Begriffsschrift]...only that part of judgments which affects the possible inferences is taken into consideration. Whatever is needed for a correct inference is fully expressed; what is not needed is...not."

III. Ingredient vs. Free-Standing Content: Semantically Separating Content from Force

Once our attention has been directed at the significance of applying a classifying concept—downstream, at the consequences of applying it, rather than just upstream, at the repeatable it discriminates, the grouping it institutes—so that *mere* classification is properly distinguished from *descriptive* classification, the necessity of distinguishing different *kinds* of consequence becomes apparent. One distinction in the vicinity, which has already been mentioned in passing, is that between *practical* and *theoretical* (or, better, *cognitive*) consequences of application of a concept. The significance of classifying an object by responding to it one way rather than another may be to make it appropriate to *do* something else with or to it—to keep it, examine it, or throw it away, to flee or pursue or consume it, for example. This is still a matter of inference; in this case, it is *practical* inferences that are at issue. But an initial classification may also contribute to further classifications: that what is in my hand falls under both the classifications raspberry and red makes it appropriate to classify it also as ripe—which in turn has practical consequences of application (such as, under the right circumstances "falling to without further ado and eating it up," as Hegel says in another connection) that neither of the other classifications has individually. Important as the distinction between practical and cognitive inferential consequences is, in the present context there is reason to emphasize a different one.

Discursive intentional phenomena (and their associated concepts), such as assertion, inference, judgment, experience, representation, perception, action, endorsement, and imagination typically involve what Sellars calls "the notorious 'ing'/'ed' ambiguity." For under these headings we may be talking about the *act* of asserting, inferring, judging, experiencing, representing, perceiving, doing, endorsing, and imagining, or we may be talking about the *content* that is asserted, inferred, judged, experienced, represented, perceived, done, endorsed, or imagined. 'Description' is one of these ambiguous terms (as is 'classification'). We ought to be aware of the distinction between the act of describing (or classifying), applying a concept, on the one hand, and the content of the description (classification, concept)—*how* things are described (classified, conceived)—on the other. And the distinction is not merely of theoretical importance for those of us thinking systematically about concept use. A distinctive level of conceptual sophistication is achieved by concept users that themselves distinguish between the contents of their concepts and their activity of applying them. So one thing we might want to know about a system being studied, a non-human animal, a prelinguistic human, an artifact we are building, is whether *it* distinguishes between the *concept* it applies and what it *does* by applying it.

We can see a basic version of the distinction between semantic content and pragmatic force as in play wherever *different* kinds of practical significance can be invested in the *same* descriptive content (different sorts of speech act or mental act performed using that content). Thus if a creature can not only say or think that the door is shut, but also ask or wonder whether the door is shut, or order or request that it be shut, we can see it as distinguishing in practice between the content being expressed and the pragmatic force being attached to it. In effect, it can use descriptive contents to do more than merely describe. But this sort of practical distinguishing of pragmatic from semantic components matters for the semantic hierarchy I am describing only when it is incorporated or reflected in the *concepts* (that is, the *contents*) a creature can deploy. The capacity to attach different sorts of pragmatic force to the same semantic content is not sufficient for *this* advance in structural semantic complexity. (Whether it is a necessary condition is a question I will not address—though I am inclined to think that in principle the answer is ‘No’.)

For the inferential consequences of applying a classificatory concept, when doing that is describing and not merely labeling, can be either *semantic* consequences, which turn on the *content* of the concept being applied, or *pragmatic* consequences, which turn on the *act* one is performing in applying it. Suppose John issues an observation report: “The traffic light is red.” You may infer that it is operating and illuminated, and that traffic ought to stop in the direction it governs. You may also infer that John has a visually unobstructed line of sight to the light, notices what color it is, and believes that it is red. Unlike the former inferences, these are not inferences from what John *said*, from the *content* of his utterance, from the concepts he has applied. They are inferences from his *saying* it, from the pragmatic force or significance of his *uttering* it, from the fact of his *applying* those concepts. For what he has *said*, that the traffic light is red, could be true even if John had not been in a position to notice it or form any beliefs about it. Nothing about John follows just from the color of the traffic light.[7]

It can be controversial whether a particular consequence follows from how something is described or from describing it that way, that is, whether that consequence is part of the descriptive content of an expression, the concept applied, or stems rather from the force of using the expression, from applying the concept. A famous example is expressivist theories of evaluative terms such as ‘good’. In their most extreme form, they claim that these terms have no descriptive content. *All* their consequences stem from what one is doing in using them: commending, endorsing, or approving. In his lapidary article “Ascriptivism,”[8] Peter Geach asks what the rules governing this move are. He offers the archaic term ‘macarize’, meaning to characterize someone as happy. Should we say that in apparently describing someone as happy we are not really describing anyone, but rather performing the distinctive speech act of macarizing? But why not then discern a distinctive speech act for *any* apparently descriptive term?

---

[7] One might think that a similar distinction could be made concerning a parrot that merely reliably responsively discriminated red things by squawking “That’s red.” For when he does that, one might infer that there was something red there (since he is reliable), and one might also infer that the light was good and his line of sight unobstructed. So both sorts of inference seem possible in this case. But it would be a mistake to describe the situation in these terms. The squawk is a label, not a description. We infer from the parrot’s producing it that there is something red, because the two sorts of events are reliably correlated, just as we would from the activation of a photocell tuned to detect the right electromagnetic frequencies. By contrast, John offers *testimony*. What he says is usable as a premise in our own inferences, not just the fact that his saying it is reliably correlated with the situation he (but not the parrot) reports (though they both respond to it).

[8] *The Philosophical Review*, Vol. 69, No. 2, 221-225. Apr., 1960.

What is wanted is a criterion for distinguishing semantic from pragmatic consequences, those that stem from the content of the concept being applied from those that stem from what we are doing in applying that concept (using an expression to perform a speech act). Geach finds one in Frege, who in turn was developing a point made already by Kant.[9] The logical tradition Kant inherited was built around the classificatory theory of consciousness we began by considering. Judgment was understood as classification or predication: paradigmatically, *of* something particular *as* something general. But we have put ourselves in a position to ask: is this intended as a model of how judgeable contents are constructed, or of what one is doing in judging? Kant saw, as Frege would see after him, that the phenomenon of *compound* judgments shows that it can*not* play both roles. For consider the hypothetical or conditional judgment

3] If Frege is correct, then conceptual content depends on inferential consequences.

In asserting this sentence (endorsing its content), have I predicated correctness of Frege (classified him as correct)? Have I described him as correct? Have I applied the concept of correctness? If so, then predicating or classifying (or describing) is not judging. For in asserting the conditional I have *not* judged or asserted that Frege is correct. I have at most built up a judgeable content, the antecedent of the conditional, by predication. For embedding a declarative descriptive sentence as an unasserted component in a compound asserted sentence strips off the pragmatic force its free-standing, unembedded occurrence would otherwise have had. It now contributes *only* its *content* to the *content* of the compound sentence, to which alone the pragmatic force of a speech act is attached.

---

[9] I discuss this point further in the first lecture of "Animating Ideas of Idealism" [op.cit.].

This means that embedding simpler sentences as components of compound sentences—paradigmatically, embedding them as antecedents of conditionals—is the way to discriminate consequences that derive from the *content* of a sentence from consequences that derive from the *act* of asserting or endorsing it. We can tell that 'happy' *does* express descriptive content, and is *not* simply an indicator that some utterance has the pragmatic force or significance of macarizing, because we *can* say things like:

4] If she is happy, then John should be glad.

For in asserting that, one does *not* macarize anyone. So the consequence, that John should be glad, must be due to the descriptive content of the antecedent, not to its force.
Similarly, Geach argues that the fact that we can say things like:

5] If being trustworthy is good, then you have reason to be trustworthy, shows that ‘good’ *does* have descriptive content.[10] Notice that this same test appropriately discriminates the different descriptive contents of the claims:

6] Labeling is not describing,

and

7] I believe that labeling is not describing.

For the two do not behave the same way as antecedents of conditionals. The stuttering inference

8] If labeling is not describing, then labeling is not describing,

is as solid an inference as one could ask for. The corresponding conditional

9] If I believe that labeling is not describing, then labeling is not describing,

requires a good deal more faith to endorse. And in the same way, the embedding test distinguishes [1] and [2] above. In each case it tells us, properly, that different descriptive contents are involved.

---

[10] Of course, contemporary expressivists such as Gibbard and Blackburn (who are distinguished from emotivist predecessors such as C.L. Stevenson precisely by their appreciation of the force of the Frege-Geach argument) argue that it need not follow that the right way to understand that descriptive content is not by tracing it back to the attitudes of endorsement or approval that *are* expressed by the use of the expression in free-standing, unembedded assertions.

What all this means is that any user of descriptive concepts who can also form compound sentences, paradigmatically conditionals, is in a position to distinguish what pertains to the semantic *content* of those descriptive concepts from what pertains to the *act* or pragmatic *force* of describing by applying those concepts. This capacity is a new, higher, more sophisticated level of concept use. It can be achieved *only* by looking at compound sentences in which other descriptive sentences can occur as unasserted components. For instance, it is only in such a context that one can distinguish *denial* (a kind of speech act or attitude) from *negation* (a kind of content). One who asserts [6] has *both* denied that labeling is describing, *and* negated a description. But one who asserts conditionals such as [8] and [9] has negated descriptions, but has *not* denied anything.

The modern philosophical tradition up to Frege took it for granted that there was a special attitude one could adopt towards a descriptive conceptual content, a kind of minimal force one could invest it with, that must be possible independently of and antecedently to being able to endorse that content in a judgment. This is the attitude of merely *entertaining* the description. The picture (for instance, in Descartes) was that *first* one entertained descriptive thoughts (judgeables), and *then*, by an in-principle subsequent act of will, accepted or rejected it. Frege rejects this picture. The principal—and in principle fundamental—pragmatic attitude (and hence speech act) is judging or endorsing.[11] The capacity merely to entertain a proposition (judgeable content, description) is a late-coming capacity—one that is parasitic on the capacity to endorse such contents. In fact, for Frege, the capacity to entertain (without endorsement) the proposition that *p* is just the capacity to endorse *conditionals* in which that proposition occurs as antecedent or consequent. For that is to explore its descriptive content, its inferential circumstances and consequences of application, what it follows from and what follows from it, what would make it true and what would be true if it were true, without endorsing it. This is a new kind of distanced attitude toward one's concepts and their contents—one that becomes possible only in virtue of the capacity to form compound sentences of the kind of which conditionals are the paradigm. It is a new level of cognitive achievement—not in the sense of a new kind of empirical knowledge (though conditionals can indeed codify new empirical discoveries), but of a new kind of semantic self-consciousness.

---

[11] In the first essay of "Animating Ideas of Idealism" [op.cit.] I discuss the line of thought that led Kant to give pride of place to judgment and judging.

Conditionals make possible a new sort of hypothetical thought. (Supposing that postulating a distinct attitude of supposing would enable one to do *this* work, the work of conditionals, would be making the same mistake as thinking that denial can do the work of negation.) Descriptive concepts bring empirical properties into view. Embedding those concepts in conditionals brings the contents of those concepts into view. Creatures that can do that are functioning at a higher cognitive and conceptual level than those who can only apply descriptive concepts, just as those who can do that are functioning at a higher cognitive and conceptual level than those who can only classify things by reliable responsive discrimination (that is, labeling). That fact sets a question for the different branches of cognitive science I mentioned in my introduction. Can chimps, or African grey parrots, or other non-human animals not just use concepts to describe things, but also semantically discriminate the contents of those concepts from the force of applying them, by using them not just in describing, but in conditionals, in which their contents are merely entertained and explored? At what age, and along with what other capacities, do human children learn to do so? What is required for a computer to demonstrate this level of cognitive functioning?

Conditionals are special, because they make *inferences* explicit—that is, put them into endorsable, judgeable, assertible, which is to say propositional form. And it is their role in inferences, we saw, that distinguishes descriptive concepts from mere classifying labels. But conditionals are an instance of a more general phenomenon. For we can think of them as operators, which apply to sentences to yield further sentences. As such, they bring into view a new notion of conceptual content: a new principle of assimilation, hence classification, of such contents. For we begin with the idea of sameness of content that derives from sameness of pragmatic force, attitude, or speech act. But the Frege-Geach argument shows that we can also individuate conceptual contents more finely, not just in terms of their role in free-standing utterances, but also accordingly as substituting one for another as arguments of operators (paradigmatically the conditional) does or does not yield compound sentences with the same free-standing pragmatic significance or force. Dummett calls these notions "free-standing" and "ingredient" content (or sense), respectively. Thus we might think that

10] It is nice here,

and

11] It is nice where I am,

express the same attitude, perform the same speech act, have the same pragmatic force or significance. They not only have the same circumstances of application, but the same consequences of application (and hence role as antecedents of conditionals). But we can see that they have different *ingredient* contents by seeing that they behave differently as arguments when we apply another operator to them. To use an example of Dummett's,

12] It is *always* nice here,

and

13] It is *always* nice where I am,

have very different circumstances and consequences of application, different pragmatic significances, and *do* behave differently as the antecedents of conditionals. But this difference in content, this sense of "different content" in which they patently do have different contents, is one that shows up *only* in the context of compounding operators, which apply to sentences and yield further sentences. The capacity to deploy such operators to form new conceptual (descriptive) contents from old ones accordingly ushers in a new level of cognitive and conceptual functioning.

Creatures that can not merely label, but describe are *rational*, in the minimal sense that they are able to treat one classification as providing a *reason* for or against another. If they can use conditionals, they can distinguish inferences that depend on the *content* of the concept they are applying from those that depend on what they are *doing* in classifying something as falling under that concept. But the capacity to use conditionals gives them more than just that ability. For conditionals let them *say* what is a reason for what, say *that* an inference is a good one. And for anyone who can do that, the capacity not just to *deny* that a classification is appropriate, but to use a *negation* operator to form new classificatory contents brings with it the capacity to say that two classifications (classifiers, concepts) are incompatible: that one provides a reason to withhold the other. Creatures that can use this sort of sentential compounding operator are not just *rational*, but *logical* creatures. They are capable of a distinctive kind of *conceptual self-consciousness*. For they can describe the rational relations that make their classifications into descriptions in the first place, hence be conscious or aware of them in the sense in which descriptive concepts allow them to be aware of empirical features of their world.

IV. Simple versus Complex Predicates

There is still a higher level of structural complexity of concepts and concept use. I have claimed that Frege should be credited with appreciating both of the points I have made so far: that descriptive conceptual classification beyond mere discriminative labeling depends on the inferential significance of the concepts, and that semantically distinguishing the inferential significance of the contents of concepts from that of the force of applying them depends on forming sentential compounds (paradigmatically conditionals) in which other sentences appear as components. In each of these insights Frege had predecessors. Leibniz (in his *New Essay on the Human Understanding*) had already argued the first point, against Locke. (The move from thinking of concepts exclusively as reliably differentially elicited labels to thinking of them as having to stand in the sort of inferential relations to one another necessary for them to have genuine descriptive content is characteristic of the advance from empiricism to rationalism.) And Kant, we have seen, appreciated how attention to compound sentences (including "hypotheticals") requires substantially amending the traditional classificatory theory of conceptual consciousness. The final distinction I will discuss, that between *simple* and *complex* predicates, and the corresponding kinds of concepts they express, is Frege's alone. No-one before him (and embarrassingly few even of his admirers after him) grasped this idea.

Frege's most famous achievement is transforming traditional logic by giving us a systematic way to express and control the inferential roles of *quantificationally complex* sentences. Frege could, as the whole logical tradition from Aristotle down to his time (fixated as it was on syllogisms) could not, handle iterated quantifiers. So he could, for instance, explain why

14] If someone is loved by everyone, then everyone loves someone,

is true (a conditional that codifies a correct inference), but

15] If everyone loves someone, then someone is loved by everyone,

is not. What is less appreciated is that in order to specify the inferences involving arbitrarily nested quantifiers ('some' and 'every'), he needed to introduce a new kind of predicate, and hence discern a structurally new kind of *concept*.

Our first grip on the notion of a predicate is as a *component* of sentences. In artificial languages we combine, for instance, a two-place predicate ‘P’ with two individual constants ‘a’ and ‘b’ to form the sentence ‘Pab’. Logically minded philosophers of language use this model to think about the corresponding sentences of natural languages, understanding

16] Kant admired Rousseau,

as formed by applying the two-place predicate ‘admired’ to the singular terms ‘Kant’ and ‘Rousseau’. The kind of inferences that are made explicit by *quantified conditionals*—inferences that essentially depend on the contents of the predicates involved—though, require us also to distinguish a one-place predicate, related to but distinct from this two-place one, that is exhibited by

17] Rousseau admired Rousseau,

and

18] Kant admired Kant,

but *not* by [16].

19] Someone admired himself,

that is, something of the form $\exists x[Pxx]$, follows from [17] and [18], but not from [16]. The property of **being a self-admirer** differs from that of **being an admirer** and from that of **being admired** (even though it entails both).

But there is no *part* of the sentences [17] and [18] that they share with each other that they do *not* share also with [16]. Looking just at the sub-sentential expressions out of which the sentences are built does not reveal the respect of similarity that distinguishes self-admiration from admiration in general—a respect of similarity that is crucial to understanding why the conditional

20] If someone admires himself then someone admires someone,

($\exists x[Pxx] \rightarrow \exists x \exists y[Pxy]$) expresses a good inference, while

21] If someone admires someone then someone admires himself,

($\exists x \exists y[Pxy] \rightarrow \exists x[Pxx]$) does not. For what [17] and [18] share that distinguishes them from [16] is not a *component*, but a *pattern*. More specifically, it is a pattern of cross-identification of the singular terms that two-place predicate applies to.

The repeatable expression-kind ‘admires’ is a *simple* predicate. It occurs as a component in sentences built up by concatenating it appropriately with a pair of singular terms. ‘x admires x' is a *complex* predicate.[12] A number of different complex predicates are associated with any multi-place simple predicate. So the three-place simple predicate used to form the sentence

22] John enjoys music recorded by Mark and books recommended by Bob,

generates not only a three-place complex predicate of the form Rxyz, but also two-place complex predicates of the form Rxxy, Rxyy, and Rxyx, as well as the one-place complex predicate Rxxx. The complex predicates can be thought of as patterns that can be exhibited by sentences formed using the simple predicate, or as equivalence classes of such sentences. Thus the complex self-admiration predicate can be thought of either as the *pattern*, rather than the *part*, that is common to all the sentences {"Rousseau admired Rousseau," "Kant admired Kant," "Caesar admired Caesar," "Brutus admired Brutus," "Napoleon admired Napoleon,"…}, or just as that set itself. Any member of such an equivalence class of sentences sharing a complex predicate can be turned into any other by a sequence of *substitutions* of all occurrences of one singular term by occurrences of another.

Substitution is a kind of *decomposition* of sentences (including compound ones formed using sentential operators such as conditionals). After sentences have been built up using simple components (singular terms, simple predicates, sentential operators), they can be assembled into equivalence classes (patterns can be discerned among them) by regarding some of the elements as systematically replaceable by others. This is the same procedure of noting invariance under substitution that we saw applies to the notion of free-standing content to give rise to that of ingredient content, when the operators apply only to whole sentences. Frege called what is invariant under substitution of some sentential components for others a '*function*'. A function can be applied to some arguments to yield a value, but it is not a *part* of the value it yields. (One can apply the function *capital of* to Sweden to yield the value Stockholm, but neither Sweden nor *capital of* is part of Stockholm.) He tied himself in some metaphysical knots trying to find a clear way of contrasting functions with *things* (objects). But two points emerge clearly. First, discerning the substitutional relations among different sentences sharing the same simple predicate is crucial for characterizing a wide range of inferential patterns. Second, those inferential patterns articulate the contents of a whole new class of concepts.

---

[12] This point, and the terminology of 'simple' and 'complex' predicates, is due to Dummett, in the second chapter of his monumental *Frege's Philosophy of Language* [op.cit.].

Sentential compounding already provided the means to build new simple concepts out of old ones. The Boolean connectives—conjunction, disjunction, negation, and the conditional definable in terms of them (A→B if and only if ~(A&~B))—permit the combination of simple predicates in all the ways representable by Venn diagrams, corresponding to the intersection, union, complementation, and inclusion of sets (concept extensions, represented by regions), and so the expression of new concepts formed from old ones by these operations. But there is a crucial class of new analytically *complex* concepts formable from the old ones that are *not* generable by such compounding procedures. One cannot, for instance, form the concept of a C such that for every A there is a B that stands to that C in the relation R. This is the complex one-place predicate logicians would represent as having the form $\{x: Cx \ \& \ \forall y{\in}A\exists z{\in}B[Rxz]\}$. As Frege says, such a concept cannot, as the Boolean ones can, be formed simply by putting together pieces of the boundaries of the concepts A,B, and C. The correlations of elements of these sets that concepts like these, those expressed by complex predicates, depend on, and so the inferences they are involved in, cannot be represented in Venn diagrams.

Frege showed further that it is just concepts like these that even the simplest mathematics works with. The concept of a natural number is the concept of a set every element of which has a successor. That is, for every number, there is another related to it as a successor (∀x∃y[Successor(x,y)). The decisive advance that Frege's new quantificational logic made over traditional logic is a *semantic, expressive* advance. His logical notation can, as the traditional logic could not, form *complex* predicates, and so both express a vitally important kind of concept, and logically codify the inferences that articulate its descriptive content.

Complex concepts can be thought of as formed by a four-stage process.

- First, put together simple predicates and singular terms, to form a set of sentences, say {Rab,Sbc,Tacd}.
- Then apply sentential operators to form compound sentences, say {Rab→Sbc, Sbc&Tacd}.
- Then substitute variables for some of the singular terms (individual constants), to form complex predicates, say {Rax→Sxy, Sxy&Tayz}.
- Finally, apply quantifiers to bind some of these variables, to form new complex predicates, for instance the one-place predicates (in y and z) {∃x[Rax→Sxy], ∀x∃y[Sxy&Tayz]}.

If one likes, this process can now be repeated, with the complex predicates just formed playing the role that simple predicates originally played at the first stage, yielding the new sentences

$\{\exists x[Rax \rightarrow Sxd], \forall x\exists y[Sxy\&Taya]\}$. They can then be conjoined, and the individual constant *a* substituted for to yield the further one-place complex predicate (in z) $\exists x[Rzx \rightarrow Sxd]\&\forall x\exists y[Sxy\&Tzyz]$. We can use these procedures to build to the sky, repeating these stages of concept construction as often as we like. Frege's rules tell us how to compute the inferential roles of the concepts formed at each stage, on the basis of the inferential roles of the raw materials, and the operations applied at that stage. This is the heaven of complex concept formation he opened up for us.

V. Conclusion

The result of all these considerations, which have been in play since the dawn of analytic philosophy, well over a century ago, is a four-stage *semantic* hierarchy of ever more demanding senses of "concept" and "concept use." At the bottom are concepts as reliably differentially applied, possibly learned, *labels* or classifications. Crudely behaviorist psychological theories (such as B. F. Skinner's) attempted to do all their explanatory work with responsive discriminations of this sort. At the next level, concepts as *descriptions* emerge when merely classifying concepts come to stand in *inferential*, *evidential*, *justificatory* relations to one another—when the propriety of one sort of classification has the practical significance of making others appropriate or inappropriate, in the sense of serving as *reasons* for and against them. Concepts of this sort may still all have observational uses, even though they are distinguished from labels by also having inferential ones.[13] Already at this level, the possibility exists of empirical descriptive concepts that can *only* be properly applied as the result of inferences from the applicability of others. These are *theoretical* concepts: a particularly sophisticated species of the genus of descriptive concepts.

At this second level, conceptual content first takes a distinctive *propositional* form; applications of this sort of concept are accordingly appropriately expressed using *declarative sentences*. For the propositional contents such sentences express just are whatever can play the role of premise and conclusion in *inferences*. And it is precisely being able to play those roles that distinguishes applications of descriptive concepts from applications of merely classificatory ones. Building on the capacity to use inferentially articulated descriptive concepts to make propositionally contentful judgments or claims, the capacity to form sentential *compounds*—paradigmatically *conditionals*, which make endorsements of material inferences relating descriptive concept applications propositionally explicit, and *negations*, which make endorsements of material incompatibilities relating descriptive concept applications propositionally explicit—brings with it the capacity to deploy a further, more sophisticated, kind of conceptual content: *ingredient* (as opposed to free-standing) content. Conceptual content of this sort is to be understood in terms of the contribution it makes to the content of *compound* judgments in which it occurs, and only thereby, indirectly, to the force or pragmatic significance of endorsing that content.

---

[13] A key part of the higher *inferential* grade of conceptuality (which includes the former, but transforms it) is that it is *multipremise material* inferences that one learns to draw as conclusions (=responses) now to Boolean combinations of the relatively enduring states that result from one's own responses.

Ingredient conceptual content, then, is what can be *negated,* or *conditionalized*. The distinctive sort of definiteness and determinateness characteristic of this sort of conceptual content becomes vivid when it is contrasted with contents that cannot appear in such sentential compounds, such as that expressed by pictures. My young son once complained about a park sign consisting of the silhouette of what looked like a Scottish terrier, surrounded by a red circle, with a slash through it. Familiar with the force of prohibition associated with signs of this general form, he wanted to know: "Does this mean 'No Scotties allowed'? Or 'No dogs allowed'? Or 'No animals allowed'? Or 'No pets allowed'"? Indeed. With pictures one has no way of indicating the degree of generality intended. A creature that can understand a claim like "If the red light is on, then there is a biscuit in the drawer," without disagreeing when the light is not on and no biscuit is present, or immediately looking for the biscuit regardless of how it is with the light, has learned to distinguish between the content of descriptive concepts and the force of applying them, and as a result can entertain and explore those concepts and their connections with each other without necessarily applying them in the sense of endorsing their applicability to anything present. The capacity in this way to free oneself from the bonds of the here-and-now is a distinctive kind of conceptual achievement

The first step was from merely *discriminating* classification to *rational* classification ('rational' because inferentially articulated, according to which classifications provide reasons for others). The second step is to *synthetic logical* concept formation, in which concepts are formed by logical compounding operators, paradigmatically conditionals and negation. The final step is to *analytical* concept formation, in which the sentential compounds formed at the third stage are *decomposed* by noting invariants under substitution. This is actually the same method that gave us the notion of ingredient content at the third stage of concept formation. For that metaconcept arises when we realize that two sentences that have the same pragmatic potential as free-standing, force-bearing rational classifications can nonetheless make different contributions to the content (and hence the force) of compound sentences in which they occur as unendorsed components—that is, when we notice that substituting one for the other may change the free-standing significance of asserting the compound sentence containing them. To form *complex* concepts, we must apply the same methodology to sub-sentential expressions, paradigmatically singular terms, that have multiple occurrences in those same logically *compound* sentences. Systematically assimilating sentences into various equivalence classes accordingly as they can be regarded as substitutional variants of one another is a distinctive kind of *analysis* of those compound sentences, as involving the application of concepts that were not *components* out of which they were originally constructed. Concepts formed by this sort of analysis are substantially and in principle more expressively powerful than those available at earlier stages in the hierarchy of conceptual complexity.   (They are, for instance, indispensible for even the simplest mathematics.)

This hierarchy is not a *psychological* one, but a *logical* and *semantic* one. Concepts at the higher levels of complexity presuppose those at lower levels, not because creatures of a certain kind cannot in practice, as a matter of fact, deploy the more complex kinds unless they can deploy the simpler ones, but because in principle it is structurally impossible to do so. Nothing could count as grasping or deploying the kinds of concepts that populate the upper reaches of the hierarchy without also grasping or deploying those drawn from its lower levels. The dependencies involved are not empirical, but (meta)conceptual and normative. The Fregean considerations that enforce the distinctions between and sequential arrangement of concept-kinds do not arise from studying how concept-users actually work, but from investigation of what concept use fundamentally is. They concern not how the trick (of concept use) is done, but what could in principle count as doing it—a normative, rather than an empirical issue. That is why it is philosophers who first came across this semantic hierarchical metaconceptual structure of concept-kinds.

But cognitive scientists need to know about it. For it is part of the job of the disciplines that cognitive science comprises to examine—each from its own distinctive point of view—all four grades of conceptual activity: the use of more complex and sophisticated kinds of concepts, no less than that of the simpler and less articulated sorts. The move from merely classificatory to genuinely descriptive concepts, for instance, marks a giant step forward in the phylogenetic development of sapience. I do not think we yet know what non-human creatures are capable of taking that step. Human children clearly do cross that boundary, but when, by what means, at what age or stage of development? Can non-human primates learn to use conditionals? Has anyone ever tried to teach them? The only reason to focus on that capacity, out of all the many linguistic constructions one might investigate empirically in this regard, is an appreciation of the kind of semantic self-consciousness about the rational relations among classifications (which marks the move from classification to rational description) that they make possible. Computer scientists have, to be sure, expended some significant effort in thinking about varieties of possible implementation of sentential compounding—for instance in exploring what connectionist or parallel distributed processing systems can do. But they have not in the same way appreciated the significance of the question of whether, to what extent, and how such "vehicleless" representational architectures can capture the full range of concepts expressed by complex predicates. (Those systems' lack of syntactically compositional explicit symbolic representations prohibits the standard way of expressing these concepts, for that way proceeds precisely by substitutional *de*composition of such explicit symbolic representations.) These are merely examples of potentially important questions raised by the hierarchy of conceptual complexity that cognitive scientists have by and large not been moved so much as to ask.

Why not? I think it is pretty clear that the answer is *ignorance*. Specifically, it is ignorance of the considerations, put forward already by Frege, that draw the bright semantic metaconceptual lines between different grades of concepts, and arrange them in a strict presuppositional semantic hierarchy. Any adequately trained cognitive scientist—even those working in disciplines far removed from computational linguistics—can be presumed to have at least passing familiarity with the similarly four-membered Chomsky hierarchy that lines up kinds of grammar, automaton, and *syntactic* complexity of languages in an array from most basic (finite state automata computing regular languages specifiable by the simplest sort of grammatical rules) to most sophisticated (two-stack pushdown automata computing recursively enumerable languages specifiable by unrestricted grammatical rules). But the at least equally significant *semantic* distinctions I have been retailing have not similarly become a part of the common wisdom and theoretical toolbox of cognitive science—even though they have been available for a half-century longer.

The cost of that ignorance, in questions not asked, theoretical constraints not appreciated, promising avenues of empirical research not pursued, is great. Failure to appreciate the distinctions and relations among fundamentally different kinds of concepts has led, I think, to a standing tendency systematically to overestimate the extent to which one has constructed (in AI) or discerned in development (whether by human children or non-human primates) or reverse-engineered (in psychology) what *we* users of the fanciest sorts of concepts do. That underlying ignorance is culpable. But it is not the cognitive scientists themselves who are culpable for their ignorance. The ideas in question are those that originally launched the whole enterprise of analytic philosophy. I think it is fair to say that as we philosophers have explored these ideas, we have gotten clearer about them in many respects. For one reason or another, though, we have not shared the insights we have achieved. We are culpable for having kept this treasure trove to ourselves. It is high time to be more generous in sharing these ideas.

Inferentialism, Normative Pragmatism, and Metalinguistic Expressivism

Chapter Two:

Some Strands of Wittgenstein's Normative Pragmatism, and Some Strains of his Semantic Nihilism

I. Strands of Normative Pragmatism

I first read the triumvirate of classical American pragmatists as an undergraduate, under the tutelage of Bruce Kuklick. He saw them as instituting a vibrant philosophical tradition that was visibly continued not only by C. I. Lewis, but by his students Goodman and Quine. (More controversially, but I believe, also correctly, he further saw the semantic holism Quine shared with Sellars as picking up a central strand of the idealist tradition—represented in the Golden Age by Lewis's teacher Josiah Royce—with which pragmatism had always been in conversation.) My Doktorvater Richard Rorty then made familiar to me an understanding of pragmatism sufficiently capacious to include such disparate and reciprocally unsympathetic philosophers as the early Heidegger and the later Wittgenstein, as well as Sellars, and Quine's student Davidson. I came to think of pragmatism as a house with many mansions, comprising a number of more or less closely related but distinct and separable commitments, relating various thinkers in the way Wittgenstein made famous under the rubric of "family resemblances." Excavating the conceptual antecedents of those various pragmatist views led me to see some of the most central among them as rooted firmly in the thought of the German Idealists, Kant and Hegel—as Peirce and Dewey had explicitly avowed.[14]

Among the most important of these antecedents is a thought that I take it serves as a fundamental orienting insight for the later Wittgenstein. This is the idea that intentionality is through and through a *normative* phenomenon. He understands that being in an intentional state, such as having a belief or an intention, includes having a kind of normative status. For it involves *committing* oneself as to how things are or are to be. In believing or intending one essentially makes oneself liable to normative assessments of the correctness of the belief or the success of the intention. And Wittgenstein is interested in a certain kind of puzzlement we might have about the nature of that normative significance.

> Someone says to me: "Show the children a game." I teach them gambling with dice, and the other says "I didn't mean that sort of game." Must the exclusion of the game with dice have come before his mind when he gave me the order?[15]

The thought is that the retrospective claim about what was meant, intended, ordered, or requested is quite correct: she did *not* mean that kind of game. But what, exactly, does that fact consist in? How did her request or the desire it expresses somehow reach out into the space of all the possible things I might have done with the intention of fulfilling it, to determine which would and which would not in fact fulfill it, which would and would not be correct according to the standard of assessment it sets? Whatever complaints one might have about the views that Kripke attributes to Wittgenstein in his book on rule-following, he is surely right in attributing the commitment he invokes to set up his problematic: the claim that someone meant plus by '+' has as an essential consequence that intending to use the symbol in that way *commits* or *obliges* him to applying it ("going on") in certain ways and not others, determines those as *correct* according to what he means. Intentional states are by definition contentful in a way that gives them an essentially *normative* significance. Furthermore, the normativity of intentionality is not limited to the case of *discursive* intentionality: the intentionality of states and expressions with propositional (hence conceptually articulated) contents, such as beliefs and intentions. Sub- or pre-discursive intentionality such as the mere pointing of a sign-post or the directedness of the states of a goal-seeking system also introduces a dimension of correctness and incorrectness of indication. It is important to Wittgenstein that already here we can raise a corresponding puzzlement about the relation of that intentional directedness to "the sign-post considered just as a piece of wood," that is, apart from its normative significance.

---

[14] Cf. the Introduction to my book *Perspectives on Pragmatism* [Harvard University Press, 2011], “From German Idealism to American Pragmatism—and Back.”

[15] *Philosophical Investigations* §70.

Kant was the first to appreciate the normativity of intentionality. He had the idea that what distinguishes the judgments and doings of knowers and intentional agents from the responses of merely natural creatures is that they are things the subjects of those acts and states are in a distinctive way *responsible* for. They are exercises of a distinctive kind of *authority* on the part of those knowers and agents: the authority to *commit* themselves, as to how things are or shall be. Sapience, awareness in the sense of apperception, consists in the capacity to commit oneself in this way, to make oneself liable to normative assessments as to the correctness of one's judgments, the success of one's actions. The *contents* of the intentional states of believing or intending set the normative standard for such assessments. Those contents accordingly determine how one has normatively *bound* oneself in judging or intending (endorsing the contents). Apperception in the sense Kant cares about is discursive intentionality. We can call the contents "*conceptual* contents." Concepts, accordingly, show up as "functions of judgment" in the sense that they determine what we have made ourselves *responsible for* in judging.

A number of Kant's most characteristic claims are relatively immediate corollaries of this founding insight into the normativity of discursive intentionality. The most pressing philosophical problem becomes understanding the "Verbindlichkeit," the "Gültigkeit," that is the normative binding force of judging and intending. Being a self or subject is possessing a distinctive kind of authority: the authority to bind oneself, to make oneself responsible by taking oneself to be responsible. This is the normative status that is Kantian autonomy. The minimal unit of sapient or apperceptive awareness is the judgment (rather than, as the tradition had it, the concept), for that is the minimal unit one can be responsible *for*. This is the logical primacy of the propositional, understood as the judgeable. The subjective form of the judgment is the "I think" which can accompany all of our judgments and is accordingly the emptiest of all representations. It is the explicit mark of *who* is responsible for judging (and acting). The objective form of the judgment is the "object = X" which marks what the judgment makes one responsible *to* for its correctness, that is, what it represents or is about. Kant accordingly pursues a normative understanding of representational purport in terms of a distinctive kind of responsibility of the represent*ing* to what counts thereby as represent*ed*, the authority of what is represented over representings of it. The "synthetic unity (characteristic) of apperception" is what results from rationally integrating each new commitment into the constellation of antecedent commitments, finding reasons justifying it, extracting consequences from it, and expelling commitments whose contents are incompatible with it. The contents judgeables must be understood to have are themselves to be made sense of in terms of the demands of this synthetic process: those contents must determine what is a reason for and against what other contents.

For Kant, our normative status as autonomous, our possession of the authority to make ourselves responsible, to bind ourselves by conceptual norms (either cognitively in judgment or practically in exercises of intentional agency) is simply an ontological fact about us, definitive of creatures like us. Hegel takes a large step to naturalizing this fundamental discursive normativity by treating the possession of this normative status as a social achievement. Indeed, for him, all normative statuses are understood as social statuses. (Slogan: “All transcendental constitution is social institution.”) More specifically, he understands normative *statuses*, including those corresponding to Kantian autonomy, as socially instituted by practical normative *attitudes* of reciprocal recognition. Norms are understood as implicit in social practices. This is his understanding of the Enlightenment insight that there were no normative statuses of authority or responsibility, no commitments or obligations, before or apart from our practices of taking or treating each other *as* authoritative, responsible, committed, and obliged.

These are lessons the classical American pragmatists take over from Kant and Hegel. They, too, see intentionality in all its guises as fundamentally a normative phenomenon. One of their master-ideas is to further naturalize the normativity of intentionality (both discursive and practical) by construing it as arising from the role intentional states play in the generically selectional processes whose paradigms are Darwinian evolution and individual learning (both supervised and unsupervised). These have in common the feedback-loop, Test-Operate-Test-

Exit (TOTE) structure. The pragmatists' model and emblem for the faculty of reason is neither the Enlightenment's reflectively representational mirror nor Romanticism's creatively illuminating lamp, but the flywheel governor that is the flexible instrument of control for the engines of the Industrial Revolution. The contemporary heirs of the specific pragmatist construal of the normativity of intentionality in terms of selectional processes epitomized by biological evolution are the teleosemanticists—philosophers of language and mind such as Kim Sterelny, David Papineau, and above all Ruth Millikan, whose development of this line of thought is the most original, sustained, and sophisticated.

Besides this model, both Peirce and Dewey take Hegel's *social* naturalizing of the normativity of intentionality as an important contribution to understanding the normativity characteristic of intentionality. The *social pragmatism about norms* that consists in understanding norms as implicit in social practices is a core strategy that Wittgenstein develops closer to our own time.

Looking back over this broad tradition, I think we can see that one orienting commitment running through it is to understanding discursive, apperceptive knowing *that* in terms of skillful practical knowing *how* (to use Ryle's terms). This methodological approach might be called "fundamental pragmatism." Placed in the context of Kant's normative insight, it is the methodological strategy of giving explanatory priority to norms *implicit* in *practices* or practical abilities to norms *explicit* in the form of *principles*. The converse explanatory strategy, which looks for something explicit in the form of a rule or principle behind every practical capacity deployed in cognition and agency, is what Dewey called “intellectualism,” (or “Platonism”).[16] The stage-setting for pragmatism of this sort is the notion of practical intentionality. This is the sort of skillful practical coping nonlinguistic organisms exhibit—epitomized at the high end by the efficient foraging strategies of orangutans and the stalking exploits of apex predators, but discernible at the low end even in the TOTE-based behavior of radar-guided missiles.

Nonlinguistic animals are already in a distinctive way oriented to or directed at (“onto”) the environing objects in their world that play significant roles in their lives. In its most basic form, fundamental pragmatism seeks to situate discursive intentionality within the larger field of this sort of practical intentionality. This project can take the form of exhibiting discursive intentionality as a kind of practical intentionality: a species of that genus. Or it can take the form of trying to show how discursively intentional abilities can arise out of more primitive sorts of skillful doing. A particularly strong form of the fundamental pragmatist program aims at exhibiting discursive practices and abilities as the results of recruiting and deploying practical abilities each of which can separately be exhibited by *non*discursive, merely practically intentional creatures. At its (implausible) limit, it takes the form of what I have called “pragmatic AI”: the attempt to show how discursive abilities can be *algorithmically elaborated* from a set of primitive abilities that are nondiscursive in the sense that each can be exhibited by creatures exhibiting only nondiscursive practical intentionality.[17]

---

[16] Recent examples are to be found in various programs in cognitive science. Hubert Dreyfus’s critique of the classical Newell-Simon program of artificial intelligence is a pragmatist assault on its intellectualism. Sophisticated intellectualist pushback against this sort of pragmatism can be found in Jason Stanley’s *Know How* [Oxford University Press, 2013].

[17] In Chapter Four of *Between Saying and Doing: Towards an Analytic Pragmatism* [Oxford University Press, 2008].

Another way of working out the overarching thesis of fundamental pragmatism concerns how the difference between practical and discursive intentionality is conceived. The classical American pragmatists saw the Enlightenment, including Kant, as having retained a spark of divinity in the form of our discursive capacities as knowers and intentional agents, by drawing a bright line between rational creatures and merely natural ones. Those thinkers accordingly showed up to Dewey, for instance, as having only insufficiently and incompletely succeeded in disenchanting, demystifying, and naturalizing us. Their pragmatism was a strategy for erasing the rationalist saltation encouraged by that bright line, by exhibiting the continuity (thought of as guaranteed by evolution) between our discursive abilities and the abilities of our nondiscursive relatives and ancestors. This is a way of putting meat on the bones of fundamental pragmatism, assimilating discursive to practical intentionality not as a species of a genus but as one extreme of a single dimension. Peirce's master idea of habits selected and retained as the genus of which both evolution and learning are species made possible the naturalistic construal of a cognitive continuum that runs from the skillful coping of the competent predator, through the practical intelligence of primitive hominids, to the traditional practices and common sense of civilized humans, all the way to the most sophisticated theorizing of contemporary scientists. A cognate aim and strategy is evidently one of those served by Wittgenstein's employment of the toy social practices he calls "Sprachspiele." Features of sophisticated discursive practices that we find particularly philosophically puzzling are to be illuminated by showing analogous features of extremely simple simple practices that could plausibly be learned by otherwise non-language-using hominins.

Fundamental pragmatism addresses relations between what one must *do* to count as engaging in or exhibiting discursive intentionality and what one must *do* to count as engaging in or exhibiting practical intentionality. In one sense or another, it claims, the former is to be understood in terms of the latter. It is in a broad sense a methodological commitment regarding the explanatory priority of the *pragmatics* of more basic practical intentionality to the pragmatics of more sophisticated discursive intentionality. Both sorts of intentionality admit an act/content, 'ing'/'ed' distinction (between a doing and what is done, a perceiving and what is perceived…), and fundamental pragmatism stays resolutely on the 'ing' side ("know*ing* how"/"know*ing* that"). At least in the case of discursive intentionality, this distinction between what one does in using a linguistic expression, or the functional role played by an intentional state, on the one hand, and its content (what is specified by the "that"-clause expressing what is known, believed, or intended), on the other, takes the form of a Fregean distinction between pragmatic force and semantic content. Another strand of pragmatism concerns the relations between these two dimensions of discursive intentionality.

For it is also a basic pragmatist idea that pragmatics, as the study of the practical *use* of expressions, or the relation of intentional states to what one goes on to do, should have a certain sort of explanatory pride of place over the theory of content: that *semantics* should answer to pragmatics. The pragmatist approach to the relations between force and content insists that the *point* of talking about meaning or content at all is the help doing so can offer to the principal enterprise of understanding what we do: proprieties normatively governing the use of expressions and the role of intentional states in providing the reasons according to which actions are normatively assessed. The conceptual, paradigmatically propositional, contents expressed by declarative sentences and invoked to specify the contents of discursive intentional states such as judgments, beliefs, and intentions are construed as theoretical auxiliaries, postulated to explain normative features of the use of sentences and the actions made intelligible by appeal to intentional states. Commitment to such an order of explanation is visible already in Kant's story, which starts with his account of what one is committing oneself to doing in judging (integrating the judgment into a larger constellation of commitments that exhibits the rational unity distinctive of apperception), and reads off of that an understanding of what sorts of judgeable contents judgings must be taken to possess in order to play their role in that process: namely contents that stand in relations of material consequence and incompatibility determining what is a reason for and against what. The same sort of envisaged order of explanation evidently animates Peirce's tradition-defining proposal to understand the meaning of a claim as consisting in "the total of all general modes of rational conduct which, conditionally upon all the possible different circumstances and desires, would ensue upon the acceptance of" it.[18]

At the methodological metalevel, pragmatism about the relations between semantics and pragmatics seeks to understand sayable, thinkable, judgeable contents (what Frege called "thoughts") in terms of what one is *doing* in asserting, thinking, judging, or treating believings as premises in reasoning, including practical reasoning about what to do. Pragmatism in this sense is a kind of functionalism about meaning or content. Within the properly pragmatist tradition, downstream from Kant's insight into the normativity of intentionality, it must take the form of a normative functionalism rather than a causal or dispositional functionalism. The system that is thought of as instituting roles and conferring meanings or contents can be taken to be an individual agent, whose intentional states are intelligible as contentful in virtue of the role they play in rationalizing (making appropriate) its behavior. Or the functional system can be taken to be a communal constellation of social discursive practices that confers meaning on performances and utterances subject to assessment according to its implicit norms, as Hegel and Dewey do. It seems clear that Wittgenstein, too, has a social practical understanding of the normativity of intentionality. One of his paradigms of practical (not yet discursive) intentionality is the sign-post. “Considered just as a piece of wood,” it is devoid of this sort of practical significance or meaning. It is only when considered in terms of the role it plays in “customs, uses, institutions,” that it is intelligible as having the significance of pointing in a direction, a significance that can be correctly or incorrectly followed. And like the classical pragmatists, he thinks this sort of practical significance is best to be understood in terms of the practical selectional processes of *learning* how to respond correctly to the sign-post: the way novices aquire the know-how to distinguish in practice correct from incorrect responses.

---

[18] Peirce, C.S., 1992 and 1999. *The Essential Peirce* (two volumes edited by the Peirce edition project), Bloomington: Indiana University Press, 1992–1999, Volume 2, p. 346.

In keeping with fundamental pragmatism, Wittgenstein seems to think that if we can just get clear about how the normativity of this sort of practical intentionality arises naturally in the context of social practices, we will no longer be puzzled by its discursive variety. Discursive intentionality is to be demystified by exhibiting it as a species of practical intentionality. The strategy is first to demystify the normativity of practical intentionality in terms of social practices—the “customs (uses, institutions)” of *PI* §199 referred to above—and then to demystify discursive intentionality by exhibiting it as continuous with, or a species of, this sort of practical intentionality.

II. Strains of Semantic Nihilism

I take it that the two principal metaconceptual axes of pragmatism are those I have introduced so far: fundamental pragmatism about the relations between practical and discursive intentionality at the level of pragmatics and methodological semantic pragmatism concerning the relations between pragmatics and semantics. There is every reason to see Wittgenstein as enlisted in the pragmatist camp as far as the first strand of pragmatist thought is concerned. If we ask further whether, within the scope of his recognition of the normativity of intentionality and his adoption of some sort of fundamental pragmatist strategy for understanding the relations between practical and discursive intentionality, Wittgenstein also endorses pragmatism about the relations between the pragmatics and semantics of discursive intentionality, the response must acknowledge a further complication. He does, I think, take it that all there is to confer semantic content on linguistic performances (and thereby also on the discursive intentional states they express) is their use, in the sense of the proprieties implicit in the discursive practices of producing and assessing such performances. And I take it he also thinks that the point of talking about propositional, conceptual, or other semantic content could only be that postulating such theoretical entities helps us to understand, or at least to codify those proprieties of use. For all that it is common to attribute to the later Wittgenstein a "use theory of meaning," his actual view seems to be rather that we should give up the notion of meaning in favor of that of use. He does not actually say "Meaning is use." What he says is things like "Don't look to the meaning, look to the use," and "Let the use of words teach you their meaning." If, as I have been doing, we use "pragmatics" in a broad sense to indicate the study of the *use* of expressions (Fregean "force" [Kraft]), and "semantics" to indicate the study of the *meaning* of expressions (Fregean "content" [Inhalt]) then it is not clear that Wittgenstein regards semantics as a legitimate enterprise. He seems to think that everything philosophers need or should want in order to understand discursive intentionality is available directly at the level of pragmatics, without the need to drill down theoretically to discern a deeper semantic level of explanation.

Wittgenstein apparently both understands meanings as theoretical entities postulated to explain use and thinks that any explanatory enterprise invoking such theoretical entities is broken-backed. He takes it that pragmatism is methodologically correct about the explanatory task meanings are postulated to perform—namely accounting for proprieties of use—but he does not endorse the semantic explanatory strategy that methodological insight invites. He thinks of the methodological point rather as telling us why we should *not* engage in semantic theorizing.[19] Like Quine, he thinks that we should give up the concept of meaning as something that can be the object of scientific theorizing about the use of linguistic expressions. Unlike Quine, he does not think that a retreat to a replacement semantic theory appealing instead to extensional metaconcepts of reference and truth conditions has any prospects of being more responsive to the underlying difficulty with theories of meaning. His skepticism about the possibility of improving our understanding of discursive practices by engaging in semantic theorizing is both more deep-rooted and more all-encompassing. It amounts to a through-going semantic nihilism.

---

[19] To adapt some Dummettian terminology, the claim is that Wittgenstein accepts a pragmatist general theory of meaning—that is, an account of what meaning consists in—but takes it that when it is properly understood that theory of meaning precludes the formulation of particular pragmatist semantic meaning theories.

In the rest of this paper I want to consider what reasons Wittgenstein has to adopt this radical anti-semantic view. I find two quite different lines of thought that Wittgenstein apparently endorses that could be called on to justify this attitude. The more familiar of them seems to me to be wrong-headed, depending on drawing a hasty and ill-considered conclusion from a sensible rejection of scientism in philosophy. The less familiar line of thought depends on a controversial but defensible and suggestive view about a central structural feature that distinguishes discursive practices from other social practices. It offers a much better rationale for in-principle skepticism about the semantic theoretical enterprise. Considerations bearing on the two different sorts of argument often appear side by side in Wittgenstein's text, so it is important to disentangle them so that their merits can be separately assessed.

Doing so is particularly important for me, since I have long been skeptical about Wittgenstein's semantic skepticism. I have been inclined to respond to the sage advice he offers not to assume that all uses of declarative sentences are in the fact-stating line of work (he doesn't think "I am in pain," is, for instance) or that all uses of singular terms should be understood as purporting to refer to particular objects ("the beetle in my box," for instance) by rebuking him for not going on to tell us what distinguishes those uses of declarative sentences that *are* in the fact-stating line of work from the rest, and what distinguishes those uses of singular terms that *do* purport to refer to particular objects. That is, I have been inclined to fault Wittgenstein for not offering a systematic theory of the core work-day practices of using sentences and terms in asserting and referring that he distinguishes from the peripheral and parasitic uses where language has "gone on holiday." One of my principal concerns in *Making It Explicit* and

Between Saying and Doing has been to offer such accounts, by developing pragmatist semantic theories that fall under the Wittgenstein-inspired rubric of "use theories of meaning."

My way of developing Kant's and Wittgenstein's insight into the normativity of intentionality in the context of fundamental pragmatism about the relations between practical and discursive intentionality and methodological semantic functionalism about the relations between pragmatics and semantics within discursive intentionality has been to articulate a *rationalist* version of pragmatism about discursive intentionality. Rather than the continuity Dewey and Wittgenstein see between discursive and nondiscursive practices and abilities, I take there to be a bright line distinguishing them. What makes something a specifically *linguistic* practice is that some performances are implicitly accorded the significance of *assertings* or *claimings*. These are by definition performances that can both serve as and stand in need of *reasons*, entitling their performers to the commitments they undertake by asserting. Playing this pragmatic role as reasons for and against other claimables means that the contents asserted (judged, believed) are *inferentially* articulated, standing in relations of material consequence (implication) and incompatibility, as premises and conclusions. This is what it is for those claimables *semantically* to take the form of propositional (hence conceptual) contents. And that is what makes the expressions themselves visible *syntactically* as having the form of declarative sentences. Asserting and inferring are accordingly understood as two sides of one coin, two features that must be displayed by any practice that includes giving and asking for reasons—which is to say, on this rationalistic line, any specifically *discursive* practice.

Practices that include the giving and asking for reasons, practices that consist in the undertaking and attributing of propositionally contentful commitments whose entitlements are always in principle in question, are qualitatively different from the more basic practices from which they arise. Once this line has been crossed, once one can explicitly say and think that things are thus-and-so, a whole new world of practical possibilities opens up. It makes possible, for instance, a distinctive kind of pragmatic and semantic self-consciousness, in which through the use of logical vocabulary such as conditionals and propositional-attitude-ascribing locutions one makes explicit essential features that otherwise remain implicit in the practice of giving and asking for reasons. On this account, by contrast to Wittgenstein's picture, language *does* have a "downtown," a core around which all of its suburbs grow and on which all of them depend.

From this point of view, one should be wary of Wittgenstein's extremely relaxed use of the term "*Sprach*spiel." The "slab" practice described in the opening paragraphs of *PI* shows up from the linguistic rationalist version of pragmatism as only a *vocal*, but not a truly *verbal* practice. What Wittgenstein carefully refers to as "calls" ([Ruf]) are not properly understood as imperatives. They are utterances that are appropriately responded to by doing one sort of thing rather than another. But to be *commands,* the claim would be, they must do so by explicitly *saying* what it is one is to do. And one cannot in this sense *say* "Bring a slab," unless one can *also say* "This is a slab." (Commanding "Shut the door," is intelligible only in the context of a practice in which one can also say "The door is shut.") That requires the iron triangle of the speech act of asserting *that* things are thus-and-so (on the side of pragmatics), the use of declarative sentences (on the side of syntax), and the propositional content asserted (on the side of semantics).

Another feature of Wittgenstein's practice that clashes with a rationalist or inferentialist pragmatism is his use of "rule" [Regel] to talk about the norms that are implicit in various practices. For rules are just what you get when you make such norms *explicit*, in the form of sentences, things one can *say*. And it is clearly one of Wittgenstein's basic lessons (an essential aspect of his fundamental pragmatism) that normativity in this explicit form must be understood to rest on and arise out of a more basic stratum of normativity that is implicit in practices ("customs, uses, institutions").

Wittgenstein's gradualist version of fundamental pragmatism denies qualitative differences between the most basic sort of practical intentionality exhibited already by nonlinguistic animals and the most sophisticated kinds of discursive intentionality—what Sellars in the final sentence of "Empiricism and the Philosophy of Mind" describes as

> the subtle and polydimensional discourse of the drawing room, the laboratory, and the study, the language of Henry and William James, of Einstein and of the philosophers who, in their efforts to break out of discourse to an *arché* beyond discourse, have provided the most curious dimension of all.

In this respect Wittgenstein's pragmatism is united with classical pragmatism against the rationalist linguistic pragmatism I espouse. In effect, like Dewey, he does not find any use for the distinction I have been employing between specifically *discursive* intentionality (sapience, apperception, characterized by *propositionally* contentful states and expressions) and *practical* intentionality of the sort exhibited by nondiscursive creatures more generally. His toy "Sprachspiele" typically, and purposefully, occupy an ambiguous middle ground. He does not see a bright line to be drawn here. But insisting on continuities of this sort does not by itself preclude engaging in pursuing a functionalist approach to semantics in a pragmatist spirit. One strain of Wittgenstein's thought that does support this stronger conclusion is his official resistance to offering, or describing what he is doing as offering, any philosophical *theories* at all, whether specifically *semantic* ones or not. That is, besides any doubts one might have about pursuing semantics, rather than being content with pragmatics, there is the fact that methodological pragmatism about semantics address the nature and rationale of semantic *theorizing*. And Wittgenstein seems hostile to the very idea of *theories* in philosophy. This is what Crispin Wright has called LW's "theoretical quietism."

Throughout the *Philosophical Investigations*, Wittgenstein works with a distinction between *describing* and *explaining*. He is concerned to insist that what philosophers ought to do is describe, not explain.

> It was true to say that our considerations could not be scientific ones…. And **we may not advance any kind of theory**. There must not be anything **hypothetical** in our considerations. **We must do away with all explanation, and description alone must take its place**. And this description gets its light, that is to say its purpose, from the philosophical problems. These are, of course, **not empirical** problems; they are solved, rather, by looking into the workings of our language, and that in such a way as to make us recognize those workings: in despite of an urge to misunderstand them. The problems are solved, not by giving new information, but by arranging what we have always known. [*PI* §109—emphasis added.]

Here explanation, theorizing, science, and empirical problems are lined up together and contrasted with description and rearrangement of familiar facts or phenomena as what is proper to philosophical inquiry. A cognate trope diagnoses the urge to misunderstand philosophical issues as rooted in the fact that “We feel as if we had to *penetrate* phenomena,” [*PI* §90] to dig down to “an essence that is hidden from us” [*PI* §92]. We tend to think: "But the words, significantly uttered, have after all not only a surface, but also the dimension of depth!" [*PI* §594], where what we are looking for in fact is to be found just in the proper description or arrangement of what already lies on the surface. We need not and should not try to look beyond the way linguistic expressions are used.

Here is one line of thought that might be invoked to justify the rejection of explanation in favor of description, the spurning of the idea of digging below the surface of phenomena to some deeper, underlying essence. One idea that seems to have been a constant throughout Wittgenstein’s philosophical development is the conviction that, as he puts it in the *Tractatus*: “Philosophy is not one of the natural sciences” [4.111]. Failure to appreciate this is an objectionable kind of *scientism* about philosophy. Rejecting the idea that philosophical problems are empirical ones, that philosophers should seek to explain things, that they should offer philosophical theories, are all to be understood as consequences of rejecting philosophical scientism. These consequences follow if one identifies science, in the science/philosophy contrast, with the method of postulating unobservable theoretical entities in order to explain observable phenomena. What is observable, what is available to be described, shows up as the surface. Explanation of those phenomena is by appeal to what is deeper, in the sense of not observable, not available to mere description of phenomena—that is, to theoretically hypothesized entities.

The thought is that the project of looking beyond or behind descriptions of the use of language (pragmatics) to explain those implicit practical proprieties by postulating meanings as theoretical entities (semantics) is assimilating philosophy to empirical natural sciences. Meanings are unobservable, theoretically postulated entities that stand to observable linguistic behavior as theoretically postulated molecules stand to the observable temperature, pressure, and volume properties of gases. As such, they are illegitimate for philosophical purposes. Only if one failed to appreciate that philosophy is not one of the natural sciences would one engage in theoretical postulation of this sort of hypothetical, because unobservable, entity.

Sometimes Wittgenstein allows a kind of philosophical explanation, in the sense of reminding us of observable, describable features of use that it is illuminating to be reminded of. It is distinguished from explanation in the objectionable sense precisely because and insofar as it remains on the surface, appealing only to what is observable in the use of expressions.

> In giving explanations I already have to use language full-blown (not some sort of preparatory, provisional one); this by itself shews that I can adduce only **exterior facts about language**. Yes, but then how can these explanations satisfy us?—Well, your very questions were framed in this language; they had to be expressed in this language, if there was anything to ask! And your scruples are misunderstandings. Your questions refer to words; so I have to talk about words. You say: the point isn't the word, but its meaning, and **you think of the meaning**

> **as a thing of the same kind as the word, though also different from the word**.
>
> Here the word, there the meaning. The money, and the cow that you can buy with it. (But contrast: money, and its use.) [*PI* §120—emphasis added.]

The "exterior facts about language" are "exterior" precisely in being on the observable, describable "surface." Wittgenstein's interlocutor here expresses that desire to penetrate to further, unobservable depths that Wittgenstein has elsewhere diagnosed as the source of philosophical misunderstandings. What there is is the use of words. Thinking of meanings as like words is thinking of them as entities. They are different from words in that they are *postulated*, merely *hypothetical* or *theoretical* entities. This is the status Wittgenstein is objecting to. To think of money as something standing behind its use, as a kind of value that is expressed in the use of money is to fetishize it, to reify it. It just is its use. To be sure, there is a difference between a mere piece of paper and money—and the difference is the use. (Compare: the sign-post considered just as a piece of wood.)

> "But the words, significantly uttered, have after all not only a surface, but also the dimension of depth!" After all, it just is the case that something different takes place when they are uttered significantly from when they are merely uttered. [*PI* §594]

There is a difference between a noise and the use of a word. The latter is meaningful. But that is to say: it is used. To postulate meanings as entities to be appealed to in explaining those proprieties of use is to address a philosophical question with the postulational explanatory tools of the empirical sciences.

I do think this line of thought is present in Wittgenstein's text. But I do not think it is what ultimately justifies his semantic nihilism, his principled skepticism about the possibility of semantic theorizing. It is good that that conclusion does not rest principally on this argument, because I take it that the argument from scientism is a bad argument. It overlooks substantial differences between the empirical theories of the natural sciences and philosophical semantic theories. And it mislocates the difficulties and challenges of the latter. Further, this dubious line of thought is interwoven with another, much weightier set of considerations. Understanding either strand of thought requires disentangling them.

The principal objection to the first line of thought is that a proper rejection of scientism about philosophy—that is, acceptance of the claim that philosophy is not an empirical natural science—does not require or support the methodological prohibition of appealing to hypothesized or postulated theoretical entities in philosophical accounts of discursive practice. Such a prohibition amounts to precluding semantics by enforcing *instrumentalism* in discursive pragmatics. And instrumentalism is no better a doctrine applied to our understanding of linguistic practice than it is for our understanding of other phenomena. Sellars makes this point in "Empiricism and the Philosophy of Mind" to begin with by distinguishing between "logical" and "philosophical" behaviorism. The two agree that the point of using mentalistic or psychological vocabulary is ultimately to make sense of the behavior of intelligent creatures: ultimately, knowers and agents. By "logical behaviorism" he has in mind the sort of Wittgenstein-inspired view Ryle develops in *The Concept of Mind*. On such a conception, not only is the target of mentalistic explanations behavior describable in non- or pre-mentalistic vocabulary, but every item appealed to in such explanations must be explicitly definable also in terms of logical constructions from items specifiable in such non- or pre-mentalistic vocabularies. The principal tool Ryle appeals to in such definitions or constructions is, of course, dispositions to behavior, codified in subjunctively robust conditionals and other alethic modal constructions.

"Philosophical behaviorism," by contrast, exploits the analogy: Behavioral vocabulary stands to mentalistic or psychological vocabulary as observational vocabulary stands to theoretical vocabulary. One should no more insist on being able to define, reduce, or construct mentalistic or psychological vocabulary in behaviorist terms than one should insist in general on being able to define, reduce, or construct theoretical vocabulary in purely observational terms—and for the same reason. Sellars's "Myth of Jones" (the constructive myth of the second half of *EPM*, paralleling the diagnostic "Myth of the Given" of the first half) offers an account of the "grammar" of thought- and sensation-talk as intelligible as introduced initially to explain discursive abilities. Although Wittgenstein is never explicitly mentioned, it seems clear that at least some of his remarks—the ones that gave aid and comfort to Ryle—are in the target area of the claim that philosophical behaviorism, with its invocation of theoretical entities not definable in behavioristic terms, is all the behaviorism philosophers need or should want.

Sellars buttresses this diagnosis with an account of the mistake he takes to be the basis for imposing the methodological strictures of logical behaviorism rather than indulging in the latitudinarian postulational method of philosophical behaviorism. It is the mistake made by instrumentalists, as opposed to realists about theoretical entities. That mistake is to think of the difference between observable and theoretically postulated entities as an *ontological* difference:

a difference between two different kinds of things. On this view, what is observable is solid, substantial, and real, while what is merely hypothesized or postulated is abstract, the product of conjecture, imagination, or whimsy. The difference in question is that between a castle whose walls will actually shield one from one's enemies and a castle in the air one merely dreams about. For the instrumentalist there is at least a genuine general question as to whether one should believe in the existence or reality of such things at all. In fact, Sellars argues, the distinction between observable and theoretical entities is not an ontological one at all. It is a methodological or epistemological difference. Theoretical entities are not a kind of thing. They are things that are known in a certain way. One way of knowing about things is inferential: drawing conclusions about them from other premises. Another is noninferential: by immediate observation. Observable things can be known about both ways. (It is part of the dismantling of the Myth of the Given to argue that the idea of something that can in principle *only* be known about *non*-inferentially, through observation, is a radical mistake.) We can draw conclusions about the (paradigmatically observable) shapes and colors of things—for instance from other, directly observed states of affairs. Purely theoretical objects and properties, by contrast, are those that are *only* knowable inferentially. Theoretical expressions do not have noninferential, observational uses.

One argument for the conclusion that this difference is methodological and not ontological is that the line between what can be observed and what we only have inferential access to is relative to a given stage in the development of our practices, and can change as those practices evolve. When the dwarf planet Pluto in the Kuiper belt was first thought about, our only epistemic access to it was inferential, by drawing conclusions from perturbations in the orbit of Neptune. It was at that point a purely theoretical object. When in 1930 Clyde Tombaugh first accurately aimed a sufficiently powerful telescope at the region of space the dwarf planet was hypothesized to occupy, it changed status from purely theoretically to observable. This was the "discovery" of Pluto. But the object didn't change. Only its relation to us did. What we could only find out about by one means became something we could also find out about in another way. Just so, in Sellars's Myth of Jones, thoughts and sensations, first postulated theoretically to explain regularities of behavior, become something those who have them can report noninferentially. Rorty then completes this just-so story about the origin of the Cartesian mind (in a way Wittgenstein would surely have applauded) by describing how a shift in social practices of attributing authority to the reporting performances of the subjects of thoughts could engender the incorrigibility of such reports—not because of their privileged ontological status, but because of a change in the social practices that institute their normative status as authoritative in the face of various sorts of challenge. His "eliminative materialism" thought experiment considers the possibility that, having gained Cartesian minds as regions of incorrigibility by one shift in social practice, we should lose those minds by another such shift.

For many years Michael Dummett championed specifically semantic instrumentalism, under the banner of "anti-realism." He started from the idea that meaning and understanding are co-ordinate concepts. Meanings are, in the first instance, what one understands. Grasp of meaning is a kind of understanding. Conjoining this with the Wittgensteinian thought that understanding must be practically manifestable as some sort of ability, he concludes that it is illegitimate for a theory of meaning to appeal to any items that cannot be defined in terms of their behavioral manifestations. The good thought is the idea, cognate to that common to logical and philosophical behaviorism, that the point of invoking meanings is to explain (proprieties of) the use of linguistic expressions. But that good thought does not, as Dummett thought, preclude explanations that appeal to items not definable in terms of the linguistic behavior they are invoked to explain. That is, it does not preclude the invocation of meanings as only inferable from specifications of linguistic behavior in a nonsemantic vocabulary, rather than definable without remainder in such a vocabulary. It does not require jettisoning the idea of truth-conditions in favor of assertibility conditions, or reference in favor of recognition conditions. There are constraints imposed by understanding truth-conditions and reference as theoretical postulates invoked to explain, for instance, the norms that practically govern fact-stating assertions of sentences and object-recognizing uses of singular terms. But those methodological norms are not so restrictive as to forbid the semantic notions appealed to inferentially outrunning what is observable at the level of non-semantically described linguistic behavior.

Quick as this rehearsal of considerations is, I hope it is clear that it sketches a colorable argument against semantic instrumentalism. But does it just amount to a flat-out denial of Wittgenstein’s claim that philosophy—at least insofar as philosophizing might include semantic theorizing (or, in Sellars’s case, theorizing in the philosophy of mind)—is not a natural or empirical science? Doesn’t it just identify philosophical explanation with scientific explanation? No. For Wittgenstein’s problematic (and Sellars’s, too) is framed by the Kant’s insight into the fundamentally *normative* character of intentionality. What is to be explained (or illuminated theoretically) is *proprieties* of discursive practice—not in the first instance empirical regularities or dispositions to linguistic behavior, but how it would be *correct* to use expressions, how they *ought* to be used. Regularities and dispositions come into the story only insofar as they affect or are affected by the semantogenic norms implicit in discursive practice. The fact that the discursive practice addressed by philosophical explanations, including those that postulate purely theoretical entities such as meanings or propositional contents, are to be described in an essentially *normative* vocabulary enforces a major, principled distinction between this sort of theoretical explanation and those pursued by natural sciences. Indeed, in this connection, the Kantian distinction between the normative and the natural is the one most relevant to demarcating the natural sciences by their objects. It is because Wittgensteinian "grammar," in its widest and most elastic sense, is an inquiry aimed at explaining normative phenomena that it is not a natural science. (Think in this connection of Wittgenstein's remarks in *PI* §81 about Ramsey's characterization of logic as a "normative science.")

The upshot of this argument is that rejecting scientism about philosophy—resisting the assimilation of philosophical theorizing to empirical theorizing in the natural sciences—does not provide good reasons for counting theoretical postulation of merely inferentially accessible episodes and processes as illegitimate in engendering specifically philosophical understanding. Philosophical explanation and theorizing should not be ruled out on these grounds. Is there then no point to Wittgenstein's privileging of description over explanation, his advice to stay at the observable surface rather than trying to delve theoretically into what lies hidden below that surface, in addressing potentially puzzling features of discursive practices? Is this line of thought simply mistaken? I do not think it is. For there is another set of considerations that supports a version of this methodological stricture.

III. A Better Argument: Linguistic Practice as Dynamic and Self-Transforming

I take it that Wittgenstein also takes the home language game of the concept of meaning 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 idea, 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 *systematic* 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. Here is one passage in which Wittgenstein asserts the connection between the image of family resemblances and the demand to stay on the descriptive surface rather than seeking to penetrate to further explanatory:

> In case (162) the meaning of the word "to derive" stood out clearly. But we told ourselves that this was only a quite special case of deriving; deriving in a quite special garb, which had to be stripped from it if we wanted to see the essence of deriving. So we stripped those particular coverings off; but then deriving itself disappeared.—In order to find the real artichoke, we divested it of its leaves. For certainly (162) was a special case of deriving; what is essential to deriving, however, was not hidden beneath the surface of this case, but this 'surface' was one case out of the family of cases of deriving.
>
> And in the same way we also use the word "to read" for a family of cases. And in different circumstances we apply different criteria for a person's reading.
>
> [*PI* §164]

The attempt to strip off contingent, adventitious details of one particular sort of case to penetrate to a general essence common to all yields nothing recognizable as determinately contentful. All we can do is observe the relations among a variety of cases, related like the overlapping strands making up a rope. It is easy to understand this line of thought as entailing a straightforward denial of the possibility of semantic analysis in the classical sense. But we might notice that this consideration, at least, does not speak against treating some subset of the familially related cases as paradigmatic, as defining a model to which other cases can then be related by a commentary pointing out respects of similarity and difference.

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 ^{s}essence^{s} proper-name usage could be taken to have had.[20] 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 "meaning" 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.[21]

---

[20] 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."

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 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.[22] 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. On this conception, the classical project of semantic theory is 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*.

---

[21] I am indebted for this way of thinking of Wittgenstein's point to Hans Julius Schneider's penetrating discussion in *Wittgenstein's Later Theory of Meaning: Imagination and Calculation* [Wiley-Blackwell, 2013].

[22] 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* (Oxford: Oxford University Press, 2006).

I think there is real force to this diagnosis. I suggested above that Kant's and Wittgenstein's insight into the essentially normative character of intentionality and discursive practice already makes room for a substantial distinction between natural scientific theories and explanations, on the one hand, and philosophical semantic theories and explanations on the other. Now we see Wittgenstein emphasizing another feature that distinguishes the discursive phenomena that are the object of such philosophical theorizing and explaining. A characteristic distinguishing feature of linguistic practices is their protean character, their plasticity and malleability, the way in which language constantly overflows itself, so that any established pattern of usage is immediately built on, developed, and transformed. The very act of using linguistic expressions or applying concepts transforms the content of those expressions or concepts. The way in which discursive norms incorporate and are transformed by novel contingencies arising from their usage is not itself a contingent, but a necessary feature of the practices in which they are implicit.

It is easy to see why one would see the whole enterprise of semantic theorizing as wrong-headed if one thinks that, insofar as language has an essence, that essence consists in its restless self-transformation (not coincidentally reminiscent of Nietzsche's "self-overcoming"). Any theoretical postulation of common meanings associated with expression types that has the goal of systematically deriving all the various proprieties of the use of those expressions according to uniform principles will be seen as itself inevitably doomed to immediate obsolescence as the elusive target practices overflow and evolve beyond those captured by what can only be a still, dead snapshot of a living, growing, moving process. It is an appreciation of this distinctive feature of discursive practice that should be seen as standing behind Wittgenstein's pessimism about the feasibility and advisability of philosophers engaging in semantic theorizing—not a bad instrumentalist conclusion drawn from commitment to a well-taken anti-scientism about philosophy.

And the idea that the most basic linguistic know-how is not mastery of proprieties of use that can be expressed once and for all in a fixed set of rules, but the capacity to stay afloat and find and make one's way on the surface of the raging white-water river of discursive communal practice that we always find ourselves having been thrown into (Wittgensteinian Geworfenheit) is itself a pragmatist insight. It is one that Dewey endorses and applauds. And it is a pragmatist thought that owes more to Hegel than it does to Kant. For Hegel builds his metaphysics and logic around the notion of determinate negation because he takes the normative obligation to *do* something to resolve the conflict that occurs when the result of our properly applying the concepts we have to new situations is that we (he thinks, inevitably) find ourselves with materially incompatible commitments to be the motor that drives the unceasing further determination and evolution of our concepts and their contents. The process of applying conceptual norms in judgment and intentional action is the very same process that institutes, determines, and transforms those conceptual norms.

IV. Conclusion

At this point, having sketched what I take to be a fundamental Wittgensteinian pragmatist insight, I want to close by registering a *caveat*—anticlimactic and even churlish though the gesture might be. It seems to me that one can and should both take on board that insight into the protean character of discursive practice and still engage in the enterprise of trying to give systematic theoretical shape to at least such broad categories of the use of linguistic expressions as asserting, inferring, describing, and referring. Because they are so broad and general, the perennial possibility of the eruption of new species need not disrupt the understanding we get of these activities by looking to core cases and providing local commentaries on those general models.

A significant impetus for Wittgenstein's later philosophy is recoil from the stresses on the Tractarian representationalist picture of facts as arrangements of objects that occurs when one is obliged to contort that model by postulating new, ever more outré sorts of facts to be expressed by declarative sentences whose principal uses are not easily assimilated to ordinary empirical description. One axial achievement of the *Tractatus* is its provision of a quite different, nondescriptive model of the function of specifically *logical* vocabulary. This avoided the embarrassment of Russellian logical atomism's attempt to understand negative and conditional facts on the model of arrangements of objects. But Wittgenstein came to see that the representational understanding of the assertion of declarative sentence use in terms of the description of facts about objects requires not only distinctive kinds of color facts, but legal facts, culinary facts, nautical facts, and so on—metaphysically different kinds of fact corresponding to every distinct sort of vocabulary capable of framing declarative sentences. More metaphysically puzzling are general facts, dispositional facts, probabilistic facts, semantic facts, intentional facts, normative facts, and fictional facts. Construing them on the representationalist model of arrangements of objects requires not only contortions of the notion of arrangement, but perhaps more fundamentally that of object. Hewing to this picture requires postulating exotic kinds of objects to go with singular terms that have quite different uses: universals, merely possible objects, probabilities, propositions, norms or values, and fictional characters such as Sherlock Holmes's maternal grandmother. Small wonder Wittgenstein urges us to jettison the restrictive representational model that obliges us to engage in such extravagant metaphysical extrapolations. We should reject assimilating all uses of declarative sentences to descriptive fact-stating, and reject assimilating all uses of singular terms to purporting to refer to objects.

Fair enough. But it is a long way from rejecting *this* general model and *its* postulations—now that we have seen the strains involved in applying it in discursive regions well removed from the ordinary empirical descriptions ("The frog is on the log,") that motivated the representationalist picture—to rejecting theoretical postulation in the service of generalization about discursive practice *tout court.* We can still try to say something illuminating about the what is distinctive of the core cases where declarative sentences *do* have the job of description or fact-stating, and singular terms *do* purport to pick out unique objects. It is a legitimate response to Wittgenstein's considerations to develop an alternative model to the representationalist one whose expressive limits he has led us to appreciate. We may do so in full understanding and expectation that the second model, like the first, will work reasonably well only for some regions of our practice, and will turn out to be of less and less help as we move farther away from the practices that provide its paradigm. But understanding can advance also by stitching together patchworks out of such locally helpful theories. (Mark Wilson's *Wandering Significance*, mentioned above, argues that such patchworks are an absolutely crucial form of conceptual understanding.) We understand discursive practice best by seeing which bits are best understood on one model and which on another. The aim of producing further frameworks should not be thought of as finding one that will do once and for all, everywhere. Illumination proceeds from taking many theoretical paths through the woods, and coming to appreciate which features of which phenomena stand out most clearly from which vantage-points. The counsel of wisdom here is experimental, irenic, and pluralistic: let a hundred theories blossom, let a thousand postulated entities contend. Most of what is wrong with systematic philosophical theorizing is a function of its being pursued in a Procrustean manner.[23] We blind ourselves if we take what is not smoothly reconstructable in our favored theoretical terms to be for that reason somehow illegitimate, rather than just learning a useful fact about what is and is not helpfully addressed in those terms.[24]

---

[23] In the *Afterword* to *Between Saying and Doing* I discuss further the sort of illumination one can gather from constructing alternative metaphysical idioms that aim at theoretically regimented sayings of everything that can be said.

[24] In a classic paper, Sellars sets the goal of clearing room for a view that goes beyond what he refers to as 'descriptivism' or 'factualism', a view that sees all claims as 'empirical' in a narrow sense. He says: "[O]nce the tautology 'The world is described by descriptive concepts' is freed from the idea that the business of all non-logical concepts is to describe, the way is clear to an *ungrudging* recognition that many expressions which empiricists have relegated to second-class citizenship in discourse are not *inferior*, just *different*." "Counterfactuals, Dispositions,

More specifically, the theoretical path forward that I have been recommending we try next in response to Wittgenstein's insights, both early and late, includes the following leading ideas. First, in keeping with the underlying Kant-Wittgenstein insight into the normativity of intentionality, to try to regiment a normative theoretical vocabulary for characterizing the use of linguistic expressions. I have proposed thinking of pragmatics in terms of the *commitments* interlocutors undertake, paradigmatically by making claims or assertions, and how *entitlements* to those commitments can be secured, paradigmatically by giving reasons for them. The goal is to understand the practical discursive know-how that is mastery of the use of an expression in terms of the ability practically to distinguish what someone (perhaps oneself) would be committing herself to by asserting it and what would entitle one (or preclude entitlement) to those commitments. The second idea is to use inference rather than representation as the principal semantic metaconcept, when theoretically postulating contents whose practical grasp manifests itself in the normative scorekeeping abilities specified in the pragmatics. Conceptual contents are to be understood in terms of consequential and incompatibility relations among commitments and entitlements. The inferential practices (and the implication relations governing them) are understood to be what Sellars calls "material" inferences and implications. These articulate the nonlogical contents expressed by the use of various substantive vocabularies: color vocabularies, legal vocabularies, culinary vocabularies, nautical vocabularies, and so on. (The distinctive conceptual roles played by subsentential expression-kinds such as singular terms and predicates is then adumbrated in terms of their role in material *substitution* inferences.)

---

and the Causal Modalities" in *Minnesota Studies in the Philosophy of Science, Volume II: Concepts, Theories, and the Mind-Body Problem*, ed. Herbert Feigl, Michael Scriven, and Grover Maxwell (Minneapolis: University of Minnesota Press, 1958), p.225-308. §79.

A final metatheoretical idea develops what Sellars made of what Carnap made of Wittgenstein's pathbreaking treatment of logical vocabulary in the *Tractatus*.[25] It is an approach to understanding a wide variety of vocabularies that, because of their distance along many dimensions from ordinary empirical descriptive discourse, have been thought to be particularly philosophically puzzling. These include logical vocabulary, dispositional and other alethic modal vocabulary, probabilistic vocabulary, fictional vocabulary, semantic and intentional vocabulary, and normative vocabulary (such as "commitment" and "entitlement"). The idea is that all these otherwise quite disparate kinds of vocabulary are alike in that they should be understood as in a very broad sense *metalinguistic* vocabularies. Their use is to be understood in terms of its essential expressive relations to some *other* kind of vocabulary—often, ordinary empirical descriptive vocabulary. Paradigmatically, these broadly metalinguistic relations, involving both pragmatic and semantic dimensions, include having their use be both *elaborated from* and *explicative of* features of the use of other vocabularies. That is, proprieties of the use of the metavocabulary are systematically determined by proprieties of the use of the more basic target vocabulary, and using the metavocabulary lets one *say* explicitly something significant about what one is *doing* in using the target vocabulary. In *Between Saying and Doing* I offer a botanization of such broadly metalinguistic roles vocabularies can play, showing how to recursively characterize an open-ended hierarchy of distinct expressive roles vocabularies can play with respect to other vocabularies.[26]

---

[25] I develop this thought in *From Empiricism to Expressivism: Brandom Reads Sellars* [Harvard University Press 2015], Chapter One.

[26] *Between Saying and Doing: Towards an Analytic Pragmatism* [Oxford University Press, 2008].

The motivating hope and eventual goal of taking this different theoretical path is that the combination of a systematic deontic normative pragmatic theory, an inferentialist semantic theory, and an expressivist account of logical, semantic and intentional, modal, and normative vocabularies provides a much more flexible and capacious tool for making sense of the norms that implicitly govern our multifarious linguistic practices than its pioneering representationalist forebears did, focused as they were on ordinary empirical descriptive discourse. At the least, the hope is that because this alternative approach explicitly focuses on and works best for the sorts of vocabularies least amenable to representationalist-descriptivist construal, a clearer picture will be provided by the stereoscopic vision they provide when the two approaches are laid alongside one another. Accordingly, the conclusion I think we should draw from the well-taken considerations and reminders Wittgenstein has assembled for us is not that we need *no* philosophical theories about our discursive practice, but that we need *more* of them.

Inferentialism, Normative Pragmatism, and Metalinguistic Expressivism

Chapter Three:

On the Way to a Pragmatist Theory of the Categories

1. Introduction

Several decades ago, Richard Rorty suggested that philosophical admirers of Wilfrid Sellars could be divided into two schools, defined by which of two famous passages from his masterwork "Empiricism and the Philosophy of Mind" are taken to express his most important insight. The two passages are:

> 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" (§41).

and

> [In] characterizing an episode or a state as that of *knowing*, we are not giving an empirical description of that episode or state; we are placing it in the logical space of reasons, of justifying and being able to justify what one says" (§36).[27]

The first passage, often called the "*scientia mensura*," expresses a kind of scientific naturalism. Its opening qualification is important: there are other discursive and cognitive activities besides describing and explaining. The second passage says that characterizing something as a knowing is one of them. And indeed, Sellars means that in characterizing something even as a believing or a believable, as conceptually contentful at all, one is doing something other than describing it. One is placing the item in a normative space articulated by relations of what is a reason for what. Meaning, for him, is a normative phenomenon that does not fall within the descriptive realm over which natural science is authoritative.

---

[27] In Herbert Feigl and Michael Scriven, eds., *Minnesota Studies in the Philosophy of Science*, vol. I (Minneapolis: University of Minnesota Press, 1956); reprinted in Sellars's *Science, Perception, and Reality* (London: Routledge and Kegan Paul, 1956; reissued Atascadero, Ridgeview, 1991); reprinted as a monograph, with an Introduction by Richard Rorty and a Study Guide by Robert Brandom (Cambridge, Mass.: Harvard University Press, 1997).

Rorty called those impressed by the scientific naturalism epitomized in the *scientia mensura* “right wing Sellarsians,” and those impressed by the normative nonnaturalism about semantics expressed in the other passage “left wing Sellarsians.” Acknowledging the antecedents of this usage, he used to express the hope that right wing and left wing Sellarsians would be able to discuss their disagreements more amicably and irenically than did the right wing and left wing Hegelians, who, as he put it, “eventually sorted out their differences at a six-month-long seminar called ‘the Battle of Stalingrad.’” According to this botanization, I am, like my teacher Rorty and my colleague John McDowell, a left wing Sellarsian, by contrast to such eminent and admirable right wing Sellarsians as Ruth Millikan, Jay Rosenberg, and Paul Churchland.

While I think Rorty’s way of dividing things up is helpful (there really are “41-ers” and “36-ers”), I want here to explore a different perspective on some of the same issues. I, too, will focus on two big ideas that orient Sellars’s thought. I also want to say that one of them is a good idea and the other one, on the whole, a bad idea—a structure that is in common between those who would self-identify as either right- or left-wing Sellarsians. And the one I want to reject is near and dear to the heart of the right wing. But I want, first, to situate the ideas I’ll consider in the context of Sellars's neo-Kantianism: they are his ways of working out central ideas of Kant's. Specifically, they are what Sellars makes of two fundamental ideas that are at the center of Kant's transcendental idealism: the metaconcept of *categories*, or *pure concepts of the understanding*, and the distinction between *phenomena* and *noumena*. The latter is a version of the distinction between appearance and reality, not in a light epistemological sense, but in the ontologically weighty sense that is given voice by the *scientia mensura*. I cannot say that these fall under the headings, respectively, of What Is Living and What Is Dead in Sellars's thought, since the sort of scientific naturalism he uses to interpret Kant's phenomena/noumena distinction is undoubtedly very widespread and influential in contemporary Anglophone philosophy. My aim here is to explain what I take it Sellars makes of the most promising of these Kantian ideas.

When asked what he hoped the effect of his work might be, Sellars said he would be happy if it helped usher analytic philosophy from its Humean into its Kantian phase. (*A propos* of this remark, Rorty also said, not without justice, that in these terms my own work could be seen as an effort to help clear the way from analytic philosophy's incipient Kantian phase to an eventual Hegelian one.[28]) Sellars tells us that his reading of Kant lies at the center of his work. He used that theme to structure his John Locke lectures, to the point of devoting the first lecture to presenting a version of the Transcendental Aesthetic with which Kant opens the *Critique of Pure Reason*. Those lectures, published as *Science and Metaphysics: Variations on Kantian Themes*, are Sellars's only book-length, systematic exposition of his views during his crucial middle period. The development of Kantian themes is not only self-consciously used to give that book its distinctive shape, but also implicitly determines the contours of Sellars's work as a whole. I think the best way to think about Sellars's work is as a continuation of the neo-Kantian tradition. In particular, I think he is the figure we should look to today in seeking an appropriation of Kant's theoretical philosophy that might be as fruitful as the appropriation of Kant's practical philosophy that Rawls initiated. On the theoretical side, Sellars was the greatest neo-Kantian philosopher of his generation.[29]

---

[28] In his Introduction to my Harvard University Press edition of "Empiricism and the Philosophy of Mind."

In fact, the most prominent neo-Kantians of the previous generation: C. I. Lewis and Rudolf Carnap were among the most immediate influences on Sellars's thought. Kant was the door through which Lewis found philosophy, and later, the common root to which he reverted in his attempt to reconcile what seemed right to him about the apparently antithetical views of his teachers, William James and Josiah Royce. (Had he instead been trying to synthesize Royce with Dewey, instead of James, he would have fetched up at Hegel.) In his 1929 *Mind and the World Order*, Lewis introduced as a central technical conception the notion of the sensory "Given", which Sellars would famously use (characteristically, without mentioning Lewis by name) as the paradigm of what he in *EPM* called the "Myth of the Given." (Indeed, shortly after his 1946 *An Analysis of Knowledge and Valuation*, which Sellars also clearly has in mind in *EPM*, Lewis wrote a piece addressing the question "Is The Givenness of the Given Given?" His answer was No: It is a necessary postulate of high philosophical theory, which dictates that without a sensory Given, empirical knowledge would be impossible.)

---

[29] His only rival for this accolade, I think, would be Peter Strawson, who certainly did a lot to make us realize that a reappropriation of some of Kant's theoretical philosophy might be a viable contemporary project. But I do not think of Peter Strawson's work as *systematically* neo-Kantian in the way I want to argue that Sellars's is.

In the book I argue that Sellars modeled his own Kantian "metalinguistic" treatments of modality and the ontological status of universals explicitly on ideas of Carnap. Although, like Lewis, Carnap is not explicitly mentioned in *EPM*, his presence is registered for the philosophical cognoscenti Sellars took himself to be addressing there by the use of the Carnapian term "protocol sentence" (as well as Schlick's "Konstatierung") for noninferential observations. Unlike Lewis, Carnap actually stood in the line of inheritance of classical nineteenth-century German neo-Kantianism. His teacher, Bruno Bauch, was (like Heidegger), a student of Heinrich Rickert in Freiburg—who, with the older Wilhelm Windelband, led the Southwest or Baden neo-Kantian school. In spite of these antecedents, Bauch was in many ways closer to the Marburg neo-Kantians, Hermann Cohen and Paul Natorp, in reading Kant as first and foremost a philosopher of the natural sciences, mathematics, and logic. I suppose that if one had asked Carnap in what way his own work could be seen as a continuation of the neo-Kantian tradition of his teacher, he would first have identified with this Marburg neo-Kantian understanding of Kant, and then pointed to the *logical* element of his logical empiricism—itself a development of the pathbreaking work of Frege, Bauch's friend and colleague at Jena when Carnap studied with both there—as giving a precise and modern form to the conceptual element in empirical knowledge, which deserved to be seen as a worthy successor to Kant's own version of the conceptual.

If Lewis and Carnap do not immediately spring to mind as neo-Kantians, that is because each of them gave Kant an empiricist twist, which Sellars was concerned to undo. If you thought that Kant thought that the classical empiricists' Cartesian understanding of the sensory contribution to knowledge was pretty much all right, and just needed to be supplemented by an account of the independent contribution made by a conceptual element, you might well respond to the development of the new twentieth century logic with a version of Kant that looks like Lewis's *Mind and the World Order,* and *An Analysis of Knowledge and Valuation,* and Carnap's *Aufbau* (and for that matter, Nelson Goodman's *Structure of Appearance*). That assumption about Kant's understanding of the role played by sense experience in empirical knowledge is exactly what Sellars challenges in *EPM*.

One of the consequences of his doing that is to make visible the neo-Kantian strand in analytic philosophy that Lewis and Carnap each, in his own way, represented—and which Sellars and, in our own time, John McDowell further developed. Quine was a student of both Lewis and Carnap, and the Kantian element of the common empiricism he found congenial in their thought for him drops out entirely—even though the logic remains. His Lewis and his Carnap are much more congenial to a narrative of the history of analytic philosophy initiated by Bertrand Russell and G.E. Moore, according to which the movement is given its characteristic defining shape as a recoil from Hegel (seen through the lenses of the British Idealism of the waning years of the nineteenth century). They understood enough about the Kantian basis of Hegel's thought to know that a *holus bolus* rejection of Hegel required a diagnosis of the idealist rot as having set in already with Kant. This narrative does pick out one current in the analytic river—indeed, the one that makes necessary the reappropriation of the metaconceptual resources of Kant's theoretical philosophy in the late twentieth and early twenty-first centuries. But it was never the whole story.[30] The neo-Kantian tradition comprising Lewis, Carnap, and Sellars can be thought of as an undercurrent, somewhat occluded from view by the empiricist surface.

---

[30] Paul Redding begins the process of recovering the necessary counter-narrative in the Introduction to his *Analytic Philosophy and the Return of Hegelian Thought* [Cambridge University Press, 2010].

2. Categories in Kant

Many Kantian themes run through Sellars's philosophy. My book is oriented around two master-ideas, each of which orients and ties together a number of otherwise apparently disparate aspects of his work. One is a strand of scientific naturalism, which I reject, on behalf of Sellars own better wisdom--so I claim. The one I'll focus on here is the good idea that besides concepts whose characteristic expressive job it is to describe and explain empirical goings-on, **there are concepts whose characteristic expressive job it is to make explicit necessary structural features of the discursive framework within which alone description and explanation are possible.** Failing to acknowledge and appreciate this crucial difference between the expressive roles different bits of vocabulary play is a perennial source of distinctively philosophical misunderstanding. In particular, Sellars thinks, attempting to understand concepts doing the second, framework-explicating sort of work on the model of those whose proper use is in empirical description and explanation is a fount of metaphysical and semantic confusion.[31] Among the vocabularies that play the second sort of role, Sellars includes *modal* vocabulary (not only the alethic, but also the deontic species), *semantic* vocabulary, *intentional* vocabulary, and *ontological-categorial* vocabulary (such as 'proposition', 'property' or 'universal', and 'object' or 'particular'). It is a mistake, he thinks, to understand the use of any of these sorts of vocabulary as fact-stating in the narrow sense that assimilates it to *describing* how the world is. It is a corresponding mistake to recoil from the metaphysical peculiarity and extravagance of the kinds of facts one must postulate in order to understand statements couched in these vocabularies as fact-stating in the narrow sense (e.g. normative facts, semantic facts, conditional facts, facts about abstract universals) by denying that such statements are legitimate, or even that they can be true. (Though to say that they are true is not, for Sellars, to describe them.) Both mistakes (the dogmatic metaphysical and the skeptical), though opposed to one another, stem from the common root of the *descriptivist fallacy*. That is the failure to see that some perfectly legitimate concepts do not play a narrowly descriptive role, but rather a different, explicative one with respect to the practices of description and explanation. Following Carnap, Sellars instead analyses the use of all these kinds of vocabulary as, each in its own distinctive way, "covertly metalinguistic."

---

[31] Distinguishing two broadly different kinds of *use* bits of vocabulary can play does not entail that there are two corresponding kinds of *concepts*—even in the presence of the auxiliary Sellarsian hypothesis that grasp of a concept is mastery of the use of a word. Though I suppress the distinction between these two moves in these introductory formulations, it will become important later in the story.

In opposing a Procrustean descriptivism about the expressive roles locutions can play, Sellars makes common cause with the later Wittgenstein. For Wittgenstein, too, devotes a good deal of effort and attention to warning us of the dangers of being in thrall to ("bewitched by") a descriptivist picture. We must not simply assume that the job of all declarative sentences is to state facts ("I am in pain," "It is a fact that …"), that the job of all singular terms is to pick out objects ("*I* think…," "I have a pain in my foot,"), and so on. In addition to tools for attaching, detaching, and in general re-shaping material objects (hammer and nails, saws, draw-knives…) the carpenter's tools also include plans, a foot-rule, level, pencil, and toolbelt. So, too, with discursive expressive ^{s}tools^{s}. Wittgenstein's expressive pluralism (language as a motley) certainly involves endorsement of the anti-descriptivism Sellars epitomizes by saying

> [O]nce the tautology 'The world is described by descriptive concepts' is freed from the idea that the business of all non-logical concepts is to describe, the way

> is clear to an *ungrudging* recognition that many expressions which empiricists have relegated to second-class citizenship in discourse are not *inferior*, just *different*.[32]

But Sellars differs from Wittgenstein in characterizing at least a broad class of nondescriptive vocabularies as playing generically the *same* expressive role. They are broadly *metalinguistic* locutions expressing necessary features of the framework of discursive practices that make description (and—so—explanation) possible. Of this broad binary distinction of expressive roles, with ordinary empirical descriptive vocabulary on one side and a whole range of apparently disparate vocabularies going into another class as "metalinguistic", there is, I think, no trace in Wittgenstein.[33]

The division of expressive roles that I am claiming for Sellars binds together modal, semantic, intentional, and ontological-categorial vocabulary in opposition to empirical-descriptive vocabularies traces back to Kant's idea of "pure concepts of the understanding," or categories, which play quite a different expressive role from that of ordinary empirical descriptive concepts. The expressive role of pure concepts is, roughly, to make explicit what is implicit in the use of ground-level concepts: the conditions under which alone it is possible to apply them, which is to say, use them to make judgments. Though very differently conceived, Kant's distinction is in turn rooted in the epistemological difference Hume notices and elaborates between ordinary empirical descriptive concepts and concepts expressing lawful causal-explanatory connections between them. Hume, of course, drew skeptical conclusions from the observation that claims formulated in terms of the latter sort of concept could not be justified by the same sort of means used to justify claims formulated in terms of empirical descriptive concepts.

---

[32] CDCM §79.

[33] The best candidate might be the discussion of "hinge propositions" in *On Certainty*. But the point there is, I think, different. In any case, Wittgenstein does not *generalize* the particular expressive role he is considering to anything like the extent I am claiming Sellars does.

Kant, however, looks at Newton's formulation of the best empirical understanding of his day and sees that the newly introduced concepts of force and mass are not intelligible apart from the *laws* that relate them. If we give up the claim that F equals m*a then we do not mean force and mass, but are using some at least slightly different concepts. (Galileo's geometrical version of the (late medieval) observable concept of acceleration *is* antecedently intelligible). This leads Kant to two of his deepest and most characteristic metaconceptual innovations: thinking of statements of laws formulated using alethic modal concepts as making explicit rules for reasoning with ordinary empirical descriptive concepts, and understanding the contents of such concepts as articulated by those rules of reasoning with them.

This line of thought starts by revealing the semantic presuppositions of Hume's epistemological arguments. For Hume assumes that the contents of ordinary empirical descriptive concepts are intelligible antecedently to and independently of taking them to stand to one another in rule-governed inferential relations of the sort made explicit by modal concepts. Rejecting that semantic atomism then emerges as a way of denying the intelligibility of the predicament Hume professes to find himself in: understanding ordinary empirical descriptive concepts perfectly well, but getting no grip thereby on the laws expressed by subjunctively robust rules relating them. Even though Kant took it that Hume's skeptical epistemological argument rested on a semantic *mistake*, from his point of view Hume's investigation had uncovered a crucial *semantic* difference between the expressive roles of different kinds of concepts. Once his attention had been directed to them, he set himself the task of explaining what was special about these *non*descriptive concepts.

Two features of Kant's account of the expressive role distinctive of the special class of concepts to which Hume had directed his attention are of particular importance for the story I am telling here. They are *categorial* concepts, and they are *pure* concepts. To say that they are 'categorial' in this context means that they make explicit aspects of the *form* of the conceptual as such. For Kant concepts are functions of judgment, that is, they are to be understood in terms of their role in judging. Categorial concepts express structural features of empirical descriptive judgments. What they make explicit is implicit in the capacity to make any judgments at all. This is what I meant when I said above that rather than describing how the world is, **the expressive job of these concepts is to make explicit necessary features of the framework of discursive practices within which it is possible to describe how the world is**. The paradigm here is the alethic modal concepts that articulate the subjunctively robust consequential relations among descriptive concepts.[34] It is those relations that make possible *explanations* of why one description applies because another does. That force *necessarily* equals the product of mass and acceleration means that one can *explain* the specific acceleration of a given mass by describing the force that was applied to it. (Of course, Kant also thinks that in articulating the structure of the judgeable as such, these concepts *thereby* articulate the structure of what is empirically *real*: the structure of *nature*, of the *objective world*. But this core thesis of his understanding of empirical realism within transcendental idealism is an optional additional claim, not entailed by the identification of a distinctive class of concepts as categories of the understanding.)

---

[34] Note that these concepts are *not* those Kant discusses under the heading of "Modality", but rather concern the hypothetical form of judgment.

To say that these concepts are 'pure' is to say that they are available to concept-users (judgers = those who can understand, since for Kant the understanding is the faculty of judgment) *a priori*.[35] Since what they express is implicit in any and every use of concepts to make empirical judgments, there is no *particular* such concept one must have or judgment one must make in order to be able to deploy the pure concepts of the understanding. To say that judgers can grasp these pure concepts *a priori* is *not* to say that they are *immediate* in the Cartesian sense of nonrepresentational. Precisely not. The sort of self-consciousness (awareness of structural features of the discursive as such) they make possible is mediated by those pure concepts. What was right about the Cartesian idea of the immediacy of self-consciousness is rather that these mediating concepts are available to every thinker *a priori*. Their grasp does not require grasp or deployment of any *particular* ground-level empirical concepts, but is *implicit* in the grasp or deployment of *any* such concepts. The way I will eventually recommend that we think about this distinctive *a prioricity* is that in being able to deploy ordinary empirical descriptive concepts one already knows how to do everything one needs to know how to do in order to be able to deploy the concepts that play the expressive role characteristic of concepts Kant picks out as "categorial" (as well as some that he does not).

---

[35] I take it that Kant always uses "*a priori*" and "*a posteriori*" as adverbs, modifying some some verb of cognition, paradigmatically "know".

3. Categories in Sellars

Sellars's development of Kant's idea of pure concepts of the understanding is articulated by two master ideas. First, his successor metaconception comprises concepts that are in some broad sense *metalinguistic*.[36] In pursuing this line he follows Carnap, who besides ground-level empirical descriptive vocabulary allowed metalinguistic vocabulary as also legitimate in formal languages regimented to be perspicuous. Such metalinguistic vocabulary allows the formulation of explicit rules governing the use of descriptive locutions. Ontologically classifying terms such as 'object', 'property', and 'proposition' are "quasi-syntactical" metavocabulary corresponding to overtly syntactical expressions in a proper metalanguage such as 'singular term', 'predicate', and 'declarative sentence'. They are used to formulate "L-rules", which specify the structure of the language in which empirical descriptions are to be expressed.[37] Alethic modal vocabulary is used to formulate "P-rules", which specify rules for reasoning with particular empirically contentful descriptive vocabulary. Carnap's neo-Kantianism does not extend to embracing the metaconcept of categories, which he identifies with the excesses of transcendental idealism. But in the expressions Carnap classifies as overtly or covertly metalinguistic, Sellars sees the raw materials for a more thoroughly Kantian successor conception to the idea of pure categories of the understanding.

The second strand guiding Sellars's reconceptualization of Kantian categories is his *semantic inferentialist* approach to understanding the contents of descriptive concepts. Sellars picks up on Kant's rejection of the semantic atomism characteristic of both the British empiricism of Locke and Hume that Kant was reacting to and of the logical empiricism of Carnap that Sellars was reacting to.[38] The way he works out the anti-atomist lesson he learns from Kant is in terms of the essential contribution made to the contents of ordinary empirical descriptive concepts by the inferential connections among them appealed to in *explanations* of why some descriptions apply to something in terms of other descriptions that apply to it.

---

[36] In Chapter Three I discuss the sense in which "metalinguistic" should be understood in such formulations..

[37] Chapter Seven discusses Sellars's view about this kind of locution.

> Although describing and explaining (predicting, retrodicting, understanding) are *distinguishable*, they are also, in an important sense, *inseparable*. It is only because the expressions in terms of which we describe objects, even such basic expressions as words for perceptible characteristics of molar objects, locate these objects in a space of implications, that they describe at all, rather than merely label. The descriptive and explanatory resources of language advance hand in hand.[39]

This is a rich and suggestive passage. It is worth unpacking the claims it contains. It is framed by a distinction between a weaker notion, *labeling*, and a *stronger* one, describing. By 'labeling' Sellars means discriminating, in the sense of responding differentially. A linguistic expression is used as a label if its *whole* use is specified by the *circumstances* under which it is applied—the *antecedents* of its application. We might distinguish between three kinds of labels, depending on how we think of these circumstances or antecedents. First, one could look at what stimuli as a matter of fact elicit or in fact have elicited the response that is being understood as the application of a label. Second, one could look *dispositionally*, at what stimuli *would* elicit the application of the label. Third, one could look at the circumstances in which the label is *appropriately* applied. What the three senses have in common is that they look only *up*stream, to the situations that have, would, or should prompt the use of the label. The first provides no constraint on future applications of the label—*que sera sera*—as familiar gerrymandering arguments about "going on in the same way" remind us. The second doesn't fund a notion of mistaken application. However one is disposed to apply the label is proper, as arguments summarized under the heading of "disjunctivitis" make clear. Only the third, normatively richer sense in which the semantics of a label consists in its circumstances of *appropriate* application (however the proprieties involved are understood) makes intelligible a notion of *mis*labeling.

---

[38] "Another feature of the empiricist tradition is its 'logical atomism,' according to which every basic piece of empirical knowledge is logically independent of every other. Notice that this independence concerns not only *what* is known, but the *knowing* of it. The second dimension of this 'atomism' is of particular importance for understanding Kant's rejection of empiricism…"["Towards a Theory of the Categories" §16]

[39] CDCM §108.

Sellars wants to distinguish labeling in *all* of these senses from *describing*. The idea is that since labeling of any of these sorts looks only to the *circumstances* in which the label is, would be, or should be applied, expressions used with the semantics characteristic of labels address at most one of the two fundamental aspects of the use characteristic of descriptions. The rules for the use of labels tell us something about what is (or would be or should be) in effect so described, but say nothing at all about what it is described *as*. That, Sellars thinks, depends on the *consequences* of applying one description rather than another. The semantics of genuine descriptions must look downstream, as well as upstream. It is this additional feature of their use that distinguishes descriptions from labels. (Here one might quibble verbally with Sellars's using 'label' and 'description' to describe expressions whose semantics depends on only one or on both of these dimensions of use. But it seems clear that a real semantic distinction is being marked.)

Making a further move, Sellars understands those consequences of application of descriptions as essentially involving *inferential* connections to other descriptive concepts. This is what he means by saying that what distinguishes descriptions from labels is their situation in a "space of implications." Paralleling the discussion of circumstances of application, we can think of these implications (consequences of application) as specifying what other descriptions do, would, or should *follow from* the application of the initial, perhaps responsively elicited, description. As he is thinking of things, a description (correctly) applies to a range of things (for descriptive concepts used observationally, including those that are appropriately noninferentially differentially responded to by applying the concept), which are described *by* it. And it describes them *as* something from which a further set of descriptions (correctly) follows. Crucially, these further descriptions can themselves involve applications of descriptive concepts that also have *non*-inferential (observational) circumstances of application. Descriptive concepts that have *only* inferential circumstances of application he calls 'theoretical' concepts.

In the opening sentence of the passage Sellars includes *understanding* as one of the phenomena he takes to be intricated with *description* in the way *explaining* is. Understanding a descriptive concept requires being able to place it in the "space of implications," partly in virtue of which it has the content that it does. This is in general a kind of knowing *how* rather than a kind of knowing *that*: being able to distinguish in practice the circumstances and consequences of application of the concept, when it is appropriately applied and what follows from so applying it. Grasping a concept in this sense is not an all-or-none thing. The ornithologist knows her way around inferentially in the vicinity of terms such as 'icterid' and 'passerine' much better than I do. A consequence of this way of understanding understanding is that one cannot grasp one concept without grasping many. This is Sellars's way of developing Kant's anti-atomist semantic insight.

Taking a further step (undertaking a commitment not yet obviously entailed by the ones attributed so far), Sellars also thinks that the inferences articulating the consequences of concepts used descriptively must always include *subjunctively robust* inferences. That is, the inferences making up the "space of implications" in virtue of which descriptive concepts have not only potentially atomistic circumstances of application but also non-atomistic relational consequences of application must extend to what other descriptions *would be* applicable if a given set of descriptions *were* applicable. For what Sellars means by 'explanation' is understanding the applicability of some descriptions as *explained by* the applicability of others according to just this kind of inference. This is, of course, just the sort of inferential connection that Hume's empiricist atomistic semantics for descriptive concepts, construing them as *labels*, could *not* underwrite. Sellars's conception of descriptions, as distinguished from labels, is his way of following out what he sees as Kant's anti-atomist semantic insight. *Modal* concepts make explicit these *necessary* inferential-consequential connections between descriptive concepts. They thereby perform the expressive role characteristic of Kantian categories: expressing essential features of the framework within which alone genuine description is possible.

All of this is meant to explicate what Sellars means by saying that "the descriptive and explanatory resources of language advance hand in hand." In addition to Kant's idea, Sellars here takes over Carnap's idea of understanding concepts whose paradigm is modal concepts as (in some sense) *metalinguistic*. The principal class of genuinely intelligible, nondefective nondescriptive vocabulary Carnap allows in *The Logical Syntax of Language* is syntactic metavocabulary and what he there calls "quasi-syntactic" vocabulary, which is covertly metalinguistic. For Sellars, the *rules* which modal vocabulary expresses are rules for deploying linguistic locutions. Their "rulishness" is their subjunctive robustness. Following out this line of thought, Sellars takes it that "grasp of a concept is mastery of the use of a word." He then understands the metalinguistic features in question in terms of rules of *inference*, whose paradigms are Carnap's L-rules and P-rules. His generic term for the inferences that articulate the contents of ordinary empirical descriptive concepts is "material inferences." The term is chosen to contrast with inferences that are 'formal' in the sense of depending on *logical* form. In another early essay he lays out the options he considers like this:

> ...we have been led to distinguish the following six conceptions of the status of material rules of inference:
>
> (1) Material rules are as essential to meaning (and hence to language and thought) as formal rules, contributing to the architectural detail of its structure within the flying buttresses of logical form.
>
> (2) While not essential to meaning, material rules of inference have an original authority not derived from formal rules, and play an indispensable role in our thinking on matters of fact.
>
> (3) Same as (2) save that the acknowledgment of material rules of inference is held to be a dispensable feature of thought, at best a matter of convenience.
>
> (4) Material rules of inference have a purely derivative authority, though they are genuinely rules of inference.

> (5) The sentences which raise these puzzles about material rules of inference are merely abridged formulations of logically valid inferences. (Clearly the distinction between an inference and the formulation of an inference would have to be explored).
>
> (6) Trains of thought which are said to be governed by "material rules of inference" are actually not inferences at all, but rather activated associations which mimic inference, concealing their intellectual nudity with stolen "therefores".[40]

His own position is that an expression has conceptual content conferred on it by being caught up in, playing a certain role in, material inferences:

> ...it is the first (or "rationalistic") alternative to which we are committed. According to it, material transformation rules determine the descriptive meaning of the expressions of a language within the framework provided by its logical transformation rules... In traditional language, the "content" of concepts as well as their logical "form" is determined by the rules of the Understanding.[41]

By “traditional language” here, he means Kantian language. The talk of “transformation rules” is, of course, Carnapian. In fact in this essay Sellars identifies his “material rules of inference” with Carnap’s “P-rules.” (‘Determine’ is--here, as generally--crucially ambiguous between ‘constrain’ and ‘settle’—the difference corresponding to that between what I have elsewhere called ‘weak’and ‘strong’ semantic inferentialism.)

---

[40] Sellars, "Inference and Meaning" *PPPW* pp. 265/317, reprinted in *In the Space of Reasons*.

[41] Sellars, "Inference and Meaning" *PPPW* pp. 284/336.

As already indicated, the material inferential rules that in one or another of these senses "determine the descriptive meaning of expressions" are for Sellars just the subjunctively robust, hence explanation-supporting ones. As he puts the point in the title of a long essay, he construes "Concepts as Involving Laws, and Inconceivable without Them." This is his response to Quine's implicit challenge in "Two Dogmas of Empiricism" to say what feature of their *use* distinguishes inferences determining conceptual contents from those that simply register matters of fact. Since empirical inquiry is generally required to determine what laws govern concepts such as copper, temperature, and mass, Sellars accepts the consequence that inquiry plays the role not only of determining facts but also of improving our conceptions—of teaching us more about the concepts that articulate those facts by teaching us more about what really follows from what--in a subjunctively robust, counter factual-supporting selse of "follows from."

On this way of understanding conceptual content, the *modal* concepts that express the lawfulness of connections among concepts and so underwrite subjunctively robust implications—concepts such as law, necessity, and what is expressed by the use of the subjunctive mood—have a different status from those of ordinary empirical descriptive concepts. Rather than in the first instance describing how the world is, they make explicit features of the framework that makes such description possible. Because they play this distinctive framework-explicating role, what they express must be implicitly understood by anyone who can deploy *any* ground-level descriptive concepts. As I would like to put the point, in knowing how to (being able to) use any ordinary empirical descriptive vocabulary, each interlocutor thereby already knows how to do everything she needs to know how to do, in order to be able to deploy the modal locutions that register the subjunctive robustness of the inferences that in turn determine the content of the descriptive concepts that vocabulary expresses. This is what Kant's idea that the pure concepts of the understanding are knowable *a priori* becomes when transposed into Sellars's framework.

The two lines of thought that orient Sellars's treatment of alethic modality, namely semantic inferentialism and a metalinguistic understanding of the expressive role characteristic of modal locutions, are epitomized in an early formulation:

> 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',[42]

where the rule in question is understood as a rule licensing subjunctively robust inferences. I have been filling in the claim that this overall approach to modality deserves to count as a development of Kant's notion of categories, pure concepts of the understanding, as concepts that make explicit features of the discursive framework that makes empirical description possible. Sellars himself, however, does not discuss this aspect of his work under that heading. When he talks about categories he turns instead to his nominalism about abstract entities. The central text here is "Towards a Theory of the Categories" of 1970.[43] The story he tells there begins with Aristotle's notion of categories (though he waves his hands wistfully at a discussion of its origins in Plato's *Sophist* that he feels cannot shoehorn into the paper) as ontological *summa genera*. There he opposes an unobjectionable hierarchy

> Fido is a dachshund.
>
> Fido is a dog.
>
> Fido is a brute.
>
> Fido is an animal.

---

[42] Sellars, "Language, Rules, and Behavior" footnote 2 to p. 136/296 in *PPPW*.

[43] In *Experience and Theory,* edited by L. Foster and J.W. Swanson (University of Massachusets Press, 1970), pp. 55-78. Reprinted in *Essays in Philosophy and its History* (D. Reidel, 1974).

Fido is a corporeal substance.

Fido is a substance.

To a potentially problematic one

X is a red.

X is a color.

X is a perceptual quality.

X is a quality. [44]

The next decisive move in understanding the latter hierarchy he attributes to Ockham, whom he reads as transposing the discussion into a metalinguistic key. Ockham's strategy, he tells us, is to understand

(A) Man is a species.

as

(B) ·Man· is a sortal mental term.[45]

while construing mental items as "analogous to linguistic expressions in overt speech."

This sketch sets up the transition to what Sellars makes of Kant's understanding of categories:

> What all this amounts to is that to apply Ockham's strategy to the theory of categories is to construe categories as classifications of conceptual items. This becomes, in Kant's hands, the idea that categories are the most generic functional classifications of the elements of judgments.[46]

---

[44] "Towards a Theory of the Categories" (TTC) §10-11.

[45] TTC §16.

[46] TTC §22.

At the end of this development from Aristotle through Ockham to Kant, he concludes

> [I]nstead of being *summa genera* of entities which are objects 'in the world,'…categories are *summa genera* of conceptual items.[47]

The account he goes on to expound in this essay, as well as in his other expositions of his nominalism about terms for qualities or properties, construes such terms metalinguistically, as referring to the inferential roles of the base-level concepts as used in empirical descriptions. I explain how I understand the view and the arguments on this topic in Chapter Seven of *From Empiricism to Expressivism:* "Sellars's Metalinguistic Expressive Nominalism." Without going further into that intricate view here, the point I want to make is that although Sellars does not say so, the metaconceptual role he here explicitly puts forward as a successor-concept to Kant's notion of category is generically the same as that I have argued he takes alethic modal locutions to play. It is this capacious conception I want to build upon and develop further.

4. Categories Today

The general conception of pure categorial concepts that I have been attributing to Sellars, based on the commonalities visible in his treatment of alethic modal vocabulary and of abstract ontological vocabulary, develops Kant's idea by treating some vocabularies (and the concepts they express) as "covertly metalinguistic." This Sellarsian conception represents his development of Carnap's classification of some expressions as "quasi-syntactic." The underlying insight is that some important kinds of vocabularies that are not strictly or evidently metalinguistic are used not (only) to describe things, but in ways that (also) depend on the use of *other* vocabularies—paradigmatically, empirical descriptive ones.

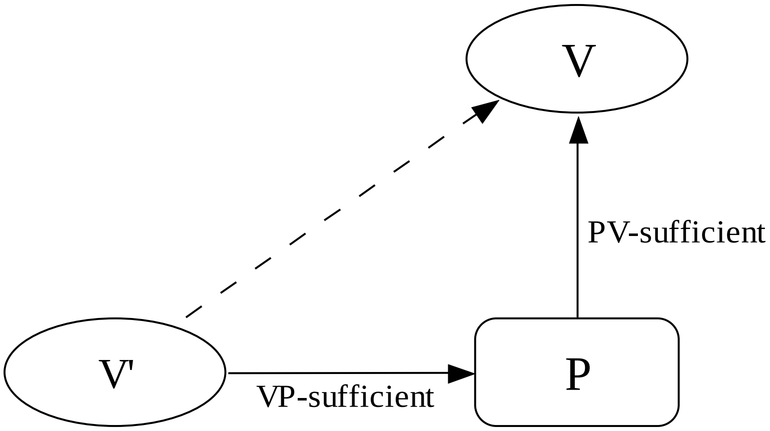
---

[47] TTC §23.

The lessons I draw from the strengths and weaknesses of Sellars's successor-conception of the "pure concepts of the Understanding" are four-fold. That is, I think he is pointing towards an expressive role characteristic of some concepts, and the vocabularies expressing them, that has four distinctive features. *First*, these concepts express what I will call "pragmatically mediated semantic relations" between vocabularies. *Second*, these concepts play the expressive role of *making explicit* essential features of the use of some other vocabulary. *Third*, the proper use of these concepts can be systematically *elaborated from* the use of that other vocabulary. *Fourth*, the features of vocabulary(concept)-use they explicate are *universal*: they are features of any and every autonomous discursive practice. I think there are concepts that play this distinctive four-fold expressive role, and that a good thing to mean today by the term "category" is metaconcepts that do so.

Carnap and Tarski introduced the expression "metalanguage" for languages that let one talk about languages, with the paradigmatic examples being syntactic and semantic metalanguages. In his earliest writings, Sellars also talks about "pragmatic metalanguages," meaning languages for talking about the *use* of expressions—rather than the syntactic or semantic properties of expressions. These were to be the languages in which we conduct what he called "pure pragmatics." During and after Sellars's most important work in the the *anni mirabiles* of 1954-63, however (possibly influenced by Carnap), he shifts to using the expression "semantics" to cover the essentially the same ground. I think that this was a step backward, and that it is one of the obstacles that prevented him from getting clear about the sense in which he wanted to claim that such locutions as alethic modal vocabulary and singular terms purporting to refer to universals ("circularity") and their kinds ("property") are "covertly metalinguistic." One vocabulary serving as a *pragmatic metavocabulary* for another is the most basic kind of pragmatically mediated semantic relation between vocabularies. It deserves to be called such because the *semantics* of the pragmatic metavocabulary depends on the *use* of the vocabulary for which it is a pragmatic metavocabulary. The relation itself is aptly called a "semantic" relation in the special case where one vocabulary is sufficient to specify practices or abilities whose exercise is sufficient to confer on another vocabulary the meanings that it expresses.

We could represent such a semantic relation, mediated by the practices of using the second vocabulary that the first vocabulary specifies, like this:[48]

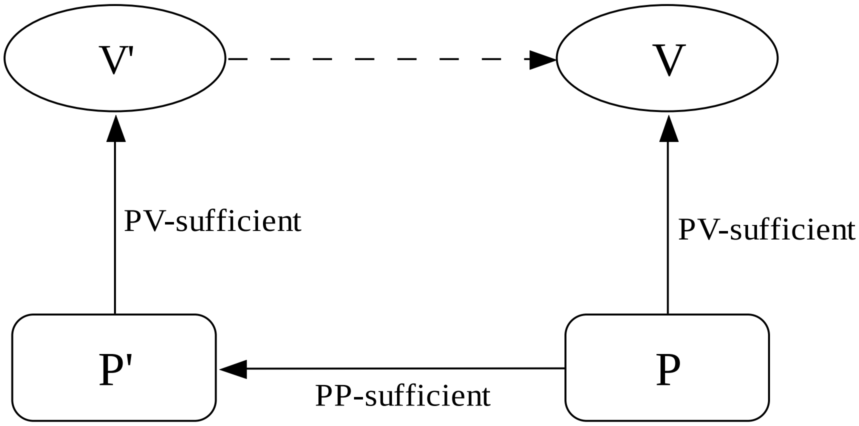


The pragmatically mediated semantic relation between vocabularies V' and V, indicated by the dashed arrow, obtains when vocabulary V' is expressively sufficient to *specify* practices-or-abilities P (that semantic fact about V' with respect to P is here called "VP-sufficiency") that are sufficient to *deploy* the vocabulary V with the meanings that it expresses when so used. In asserting that this relation between vocabularies obtains, one is claiming that if all the sentences in V' used to specify the practices-or-abilities P are true of P, then anyone engaging in those practices or exercising those abilities as specified in V' is using the expressions of V with their proper meanings. This semantic relation between what is expressible in the two vocabularies is mediated by the practices P that the first specifies and which are the use of the second. This particular pragmatically mediated semantic relation holds when the vocabulary V' allows one to *say* what one must *do* in order to *say* what can be said in the vocabulary V. In that sense V' makes *explicit* (sayable, claimable) the practices-or-abilities *implicit* in using V. This is the explicative relation I mention as the second component of the complex expressive role that I am offering as a candidate for a contemporary successor-(meta)concept to Kant's (meta)concept of category. There are other pragmatically mediated semantic relations besides being a pragmatic metavocabulary in this sense, and others are involved in the categorial expressive role. The result will still fall under the general rubric that is the first condition: being a pragmatically mediated semantic relation.

---

[48] I introduce, develop, and apply these "meaning-use diagrams" in *Between Saying and Doing: Towards an Analytic Pragmatism* [Oxford University Press, 2008].

One such further pragmatically mediated semantic relations between vocabularies holds when the practices PV-sufficient for deploying one vocabulary, though not themselves PV-sufficient for deploying a second one, can be systematically elaborated into such practices. That is, in being able to deploy the first vocabulary, one already knows how to do everything one needs to know how to do, in principle, to deploy the second. But those abilities must be suitably recruited and recombined. The paradigm here is *algorithmic* elaboration of one set of abilities into another. Thus, in the sense I am after, the capacities to do multiplication and subtraction are algorithmically elaborable into the capacity to do long division. *All* you need to learn how to do is to put together what you already know how to do in the right way—a way that can be specified by an algorithm. The diagram for this sort of pragmatically mediated semantic relation between vocabularies is:



The dotted arrow indicates the semantic relation between vocabularies V' and V. It is the relation that holds when all the relations indicated by solid arrows hold—that is, when the practices-or-abilities sufficient to deploy vocabulary V can be elaborated into practices sufficient to deploy vocabulary V'. In this case, the semantic relation in question is mediated by two sets of practices-or-abilities: those sufficient to deploy the two vocabularies.

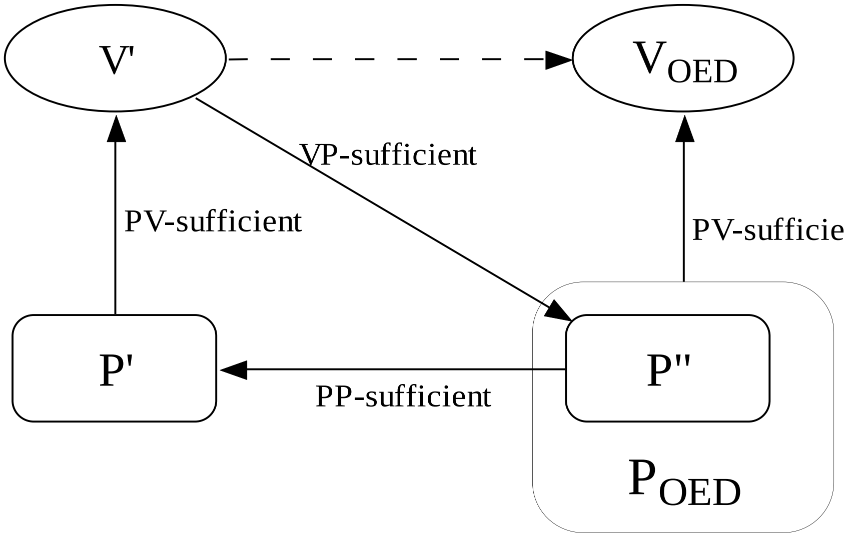
A concrete example of vocabularies standing in this pragmatically mediated semantic relation, I claim, is that of *conditionals* in relation to ordinary empirical descriptive (OED) vocabulary. For using such OED vocabulary, I claim (following Sellars following Kant), requires distinguishing in practice between materially good inferences involving descriptive predicates and ones that are not materially good. One need not be either infallible or omniscient in this regard, but unless one makes *some* such distinction, one cannot count as deploying the OED vocabulary in question. But in being able practically to distinguish (however fallibly and incompletely) between materially good and materially bad inferences, one knows how to do everything one needs to know how to do, in principle, to deploy conditionals. For conditionals can be introduced by recruiting those abilities in connection with the use of sentences formed from the old vocabulary by using the new vocabulary. On the side of circumstances of application (assertibility conditions), one must acknowledge commitment to the conditional p→q just in case one takes the inference from p to q to be a materially good one. And on the side of consequences of application, if one acknowledges commitment to the conditional p→q, then one must take the inference from p to q to be a materially good one. These rules constitute an algorithm for elaborating the ability to distinguish materially good from materially bad inference using OED vocabulary (or any other vocabulary, for that matter) into the ability appropriately to use conditionals formed from that vocabulary: to distinguish when such conditionals are assertible, and what the consequences of their assertibility is.

My idea for a successor-concept to what Sellars (with hints from Carnap) made of Kant's metaconception of pure concepts of the Understanding is that they must play *both* of these expressive roles, stand in *both* sorts of pragmatically mediated semantic relations to another vocabulary. It must be possible to *elaborate* their use from the use of the index vocabulary, and they must *explicate* the use of that index vocabulary. Speaking more loosely, we can say that such concepts are both *elaborated from* and *explicative of* the use of other concepts—in short that they are el-ex, or just LX with respect to the index vocabulary.

The fourth condition I imposed above is that the concepts in question must be *universally* LX, by which I mean that they must be LX for every autonomous discursive practice (ADP)—

every language game one could play though one played no other. That is, the practices from which their use can be elaborated and of which their use is explicative must be essential to talking or thinking at all. This universality would distinguish categorial concepts, in the sense being specified, from metaconcepts that were elaborated from and explicative of only some parasitic fragment of discourse—culinary, nautical, or theological vocabulary, for instance. I take it that any autonomous discursive practice must include the use of ordinary empirical descriptive vocabulary. If so, being LX for OED vocabulary would suffice for being *universally* LX, LX for every ADP.

Putting all these conditions together yields the following diagram of the pragmatically mediated semantic relation between vocabularies that obtains when vocabulary V' plays the expressive role of being universally LX by being elaboratable from and explicative of practices necessary for the deployment of ordinary empirical descriptive vocabulary:



The fact that the rounded rectangle labeled P", representing the practices from which vocabulary V’ is elaborated and of which it is explicative, appears inside the rounded rectangle representing practices sufficient to deploy ordinary empirical descriptive vocabulary indicates that the practices P" are a necessary part of the practices sufficient to deploy OED vocabulary, but need not comprise all such practices. Thus, distinguishing materially good from materially bad inferences involving them is necessary for deploying ordinary empirical descriptive vocabulary (rather than mere labels), but there is a lot more involved in doing so—using such vocabulary observationally, for instance. Different categorial metaconcepts can be LX for different essential features of the use of empirical descriptive vocabulary. Thus alethic modal vocabulary explicates the subjunctive robustness of the inferences explicated by conditionals. “Quasi-syntactic” abstract ontological vocabulary such as ‘property’ and ‘proposition’ explicate structural features of descriptive sentences.

Diagramming the expressive role of being LX for practices necessary to deploy OED vocabulary provides an analysis that breaks down the claim that some vocabulary plays a categorial role into its component sub-claims. To show that alethic modal vocabulary, for instance, stands in this pragmatically mediated semantic relation to ordinary empirical descriptive vocabulary one must show that there is some practices-or-abilities (in this case, to reason subjunctively or counterfactually) that are 1) a necessary component of practices-or-abilities that are 2) (PV)sufficient to deploy OED vocabulary, 3) from which one can elaborate practices-or-abilities that are 4) (PV)sufficient to deploy vocabulary (alethic modal vocabulary) 5) that is (VP)sufficient to explicate or specify the original practices-or-abilities. Although there is by design considerable elasticity in the concepts vocabulary, practices-or-abilities, and the various sufficiency and necessity relations between them, the fine structure of the distinctive expressive role in question is clearly specified.

What credentials does that expressive role have to pick out a worthy successor metaconcept to what Sellars made of Kant's categories or pure concepts of the Understanding? At the beginning of my story I introduced the idea behind the Kantian categories as the idea that besides the concepts whose principal use is in giving empirical descriptions and explanations, there are concepts whose principal use is in making explicit features of the framework that makes empirical description and explanation possible. The expressive task characteristic of concepts of this latter class is to articulate what Kant called the "transcendental conditions of experience." The concepts expressed by vocabularies that are LX for empirical descriptive vocabulary perform this defining task of concepts that are categories. As explicative of practices necessary for deploying vocabularies performing the complex expressive task of description and explanation (distinguishable only in the context of their complementary relations within a pragmatic and semantic context that necessarily involves both), this kind of vocabulary makes it possible to *say* what practitioners must be able to *do* in order to describe and explain how things empirically are. They do this by providing a pragmatic metavocabulary for describing and explaining. This is a central feature (the 'X' in 'LX') of the complex pragmatically mediated semantic relation between categorial metaconcepts and ordinary empirical descriptive vocabulary.

One feature of the concepts performing this explicative function that Kant emphasizes is that they are "*pure* concepts of the Understanding." (I take it that the "of" should be understood as expressing both the subjective and objective genitives—as in "Critique of Pure Reason." These concepts both belong to the Understanding and address it, being both discursive and metaconceptual.) To say that they are pure concepts is to say that they are graspable *a priori*.[49] The feature of the LX model that corresponds to the a prioricity of Kant's categories is that the use of LX metaconcepts can be elaborated from that of the empirical descriptive vocabularies for which they are LX. As I have put the point, in knowing how to deploy OED vocabulary, one already knows how to do everything one needs to know how to do to deploy vocabulary that is LX for it—such as alethic modal vocabulary, conditionals, and ontological classificatory vocabulary. If we take it, as per Sellars, that grasp of a concept is mastery of the use of a word, then one need not *actually* grasp concepts that are LX for descriptive vocabulary in order to deploy descriptive vocabulary. But in effect, *all* one is missing is the words for them. The circumstances and consequences of application of LX concepts can be formulated by rules that appeal only to abilities one already has in virtue of being able to use OED vocabulary. (Think of the sample rules for conditionals sketched above.) In that sense, the LX concepts are *implicit in* the descriptive concepts. It is not that one must or could grasp these concepts *before* deploying descriptive concepts. It is rather that nothing more is required to grasp them than is required to deploy descriptive concepts, and there are no particular descriptive concepts one must be able to deploy, nor any particular descriptive claims that one must endorse, in order to possess abilities sufficient to deploy the universally LX metaconcepts.

The class of concepts that are arguably universally LX (LX for every autonomous discursive practice because LX for OED vocabulary) overlaps Kant's categories in important ways—most notably in the alethic modal concepts that make explicit subjunctively robust consequential relations among descriptive concepts. But the two do not simply coincide. In

---

[49] Kant does admit also impure *a priori* principles.

Between Saying and Doing I argue that besides modal vocabulary, logical vocabulary, indexical and demonstrative vocabulary, normative vocabulary, and semantic and intentional vocabulary all should be thought of as LX for OED vocabulary. In spite of this extensional divergence, the fact that vocabulary that is LX for descriptive vocabulary in general principle shares with Kant's categories the two crucial features of being explicative of such vocabulary and being graspable *a priori* makes the idea of universally LX metaconcepts a worthy successor to Kant's breakthrough idea. The fact that Sellars's own development of this idea of Kant's takes such important steps in this direction convinces me that his version of the categories was a progressive step, and a Good Idea.

Inferentialism, Normative Pragmatism, and Metalinguistic Expressivism

Chapter Four:

From Logical Expressivism to Expressivist Logic:

Sketch of a Program and Some Implementations[50]

I. Introduction

Traditionally, two principal issues in the philosophy of logic are the demarcation question (what distinguishes specifically *logical* vocabulary?) and the correctness question (what is the *right* logic?). One of the binding-agents tying together semantic and logical inferentialism is a distinctive philosophy of logic: logical expressivism. This is the view that the expressive role that distinguishes logical vocabulary is to make explicit the inferential relations that articulate the semantic contents of the concepts expressed by the use of ordinary, nonlogical vocabulary. If one offers this logically expressivist, semantically inferentialist answer to the demarcation question, the correctness question lapses.

It is replaced by a concrete task. For each bit of vocabulary to count as logical in the expressivist sense, one must say what feature of reasoning, to begin with, with *non*logical concepts, it expresses. Instead of asking what the *right* conditional is, we ask what dimension of normative assessment of implications various conditionals make explicit. For instance, the poor, benighted, and unloved, classical two-valued conditional makes explicit the sense of "good inference" in which it is a good thing if an inference does *not* have true premises and a false conclusion. (At least we can acknowledge that implications that do *not* have at least this property are bad.) Intuitionistic conditionals in the broadest sense let us assert that there is a procedure for turning an argument for the premises of an inference into an argument for the conclusion. C.I. Lewis's hook of strict implication codifies the sense in which it is a good feature of an inference if it is *impossible* for its premises to be true and its conclusion not to be true. And so on. There can in principle be as many conditionals as there are dimensions along which we can endorse implications.

---

[50] The proof-theoretic logical systems I report on in this paper were developed as the result of many years of work in our logic working group at the University of Pittsburgh, brought to fruition by Ulf Hlobil and Dan Kaplan. We will present them, along with many more, including some by Shuhei Shimamura, in the co-authored book we are writing, *Logics of Consequence: Tools for Expressing Structure*.

In spite of its irenic neutrality concerning the correctness question, one might hope that a new approach to the philosophy of logic such as logical expressivism would not only explain features of our *old* logics but ideally also lead to *new* developments in logic itself. I think this is in fact the case, and I want here to offer a sketch of how.

II. Prelogical Structure

I take it that the task of logic is to provide mathematical tools for articulating the *structure* of reasoning. Although for good and sufficient historical reasons, the original test-bench for such tools was the codification of specifically *mathematical* reasoning, the expressive target ought to be reasoning generally, including for instance and to begin with, its more institutionalized species, such as reasoning in the empirical sciences, in law-courts, and in medical diagnosis.

We can approach the target-notion of the structure of reasoning in two stages. The first stage distinguishes what I will call the "relational structure" that governs our reasoning practices. Lewis Carroll's fable "Achilles and the Tortoise" vividly teaches us to distinguish, in John Stuart Mill's terms, "premises from which to reason" (including those codifying implication relations) from "rules in accordance with which to reason," demonstrating that we cannot forego the latter wholly in favor of the former. Gil Harman sharpens the point in his argument that there is no such thing as rules of deductive reasoning. If there were, presumably a paradigmatic one would be: If you believe *p* and you believe *if p then q*, then you should believe *q*. But that would be a terrible rule. You might have much better reasons against *q* than you have for either of the premises. In that case, you should give up one of them. He concludes that we should distinguish *relations* of *implication*, from *activities* of *inferring*. The fact that *p, if p then q,* and *not-q* are incompatible, because *p* and *if p then q* stand in the implication relation to *q,* normatively *constrains* our reasoning activity, but does not by itself *determine* what it is correct or incorrect to do.

The normative center of reasoning is the practice of assessing reasons for and against conclusions. Reasons *for* conclusions are normatively governed by relations of *consequence* or *implication*. Reasons *against* conclusions are normatively governed by relations of *incompatibility*. These relations of implication and incompatibility, which constrain normative assessment of giving reasons for and against claims, amount to the first significant level of *structure* of the *practice* of giving reasons for and against claims.

These are, in the first instance, what Sellars called "*material*" relations of implication and incompatibility. That is, they do not depend on the presence of logical vocabulary or concepts, but only on the contents of non- or pre-logical concepts. According to semantic inferentialism, these are the relations that articulate the conceptual contents expressed by the *prelogical* vocabulary that plays an essential role in formulating the premises and conclusions of inferences.

Once we have distinguished these *relations* from the practice or activity of reasoning that they normatively govern, we can ask after the *algebraic* structure of such relations. In 1930s, Tarski and Gentzen, in the founding documents of the model-theoretic and proof-theoretic traditions in the semantics of logic, though differing in many ways in their approaches (as Jarda discusses in the second half of his book), completely agree about the algebraic structure of *logical* relations of consequence and incompatibility. Logical consequence satisfies Contexted Reflexivity (or Containment), Monotonicity, and Idempotence (Gentzen's "Cut", sometimes called "Cumulative Transitivity"). In Tarski's terms: $X \subseteq Cn(X)$, $X \subseteq Y \Rightarrow Cn(X) \subseteq Cn(Y)$, and $Cn(Cn(X))=Cn(X)$. Logical incompatibility satisfies what Peregrin calls "explosion": the implication of *everything* by logically inconsistent sets. (Peregrin builds this principle in so deeply that he takes the functional expressive role of negation to be serving as an "explosion detector.")

Perhaps these are, indeed, the right principles to require of specifically *logical* relations of consequence and incompatibility. But logical expressivists must ask a prior question: **What is the structure of *material* relations of consequence and incompatibility?** This is a question the tradition has not thought about at all. But the answer one gives to it substantially shapes the logical enterprise when it is construed as expressivism does.

We can think of statements of implication and incompatibility as expressing what is *included in* a premise-set and what is *excluded by* it. In a semantic inferentialist spirit, we can say that the elements of a premise-set are its *explicit* content, and its consequences are its *implicit* content—in the literal sense of what is *implied by* it. It is reasonable to suppose that what is *explicitly* contained in a premise-set is also part of its *implicit* content. It is accordingly plausible to require that material consequence relations, no less than logical ones, be reflexive in an extended sense: if the premises explicitly contain a sentence, they also implicitly contain it, regardless of what other auxiliay premises are available. (We sometimes call this condition "Containment", thinking of Tarski's algebraic closure principle that every premise-set is a subset of its cooonsequence-set.)

Monotonicity, by contrast, is *not* a plausible constraint on *material* consequence relations. It requires that if an implication (or incompatibility) holds, then it holds no matter what additional auxiliary hypotheses are added to the premise-set. But outside of mathematics, almost all our actual reasoning is *defeasible*. This is true in everyday reasoning by auto mechanics and on computer help lines, in courts of law, and in medical diagnosis. (Indeed, the defeasibility of medical diagnoses forms the basis of the plots of every episode of “House” you have ever seen—

besides all those you haven't.) It is true of subjunctive reasoning generally. If were to I strike this dry, well-made match, it would light. But *not* if it is in a very strong magnetic field. Unless, additionally, it were in a Faraday cage, in which case it would light. But *not* if the room were evacuated of oxygen. And so on.

The idea of "laws of nature" reflects an approach to subjunctive reasoning deformed by a historically conditioned, Procrustean ideology whose shortcomings show up in the need for idealizations (criticized by Cartwright in *How the Laws of Physics Lie*) and for "physics avoidance" (diagnosed by Wilson in *Wandering Significance* on the basis of the need to invoke supposedly "higher-level" physical theories in *applying* more "fundamental" ones). Defeasibility of inference, hence nonmonotonicity of implication relations, is a structural feature not just of probative or permissive reasoning, but also of dispositive, committive reasoning. *Ceteris paribus* clauses do not magically turn nonmonotonic implications into monotonic ones. (The proper term for a Latin phrase whose recitation can do *that* is "magic spell.") The expressive function characteristic of *ceteris paribus* clauses is rather explicitly to *mark* and *acknowledge* the defeasibility, hence nonmonotonicity of an implication codified in a conditional, not to cure it by *fiat*.

The logical expressivist (including already—as I've argued elsewhere—Frege in the *Begriffsschrift*, at the dawn of modern logic) thinks of logical vocabulary as introduced to let one *say* in the logically extended object-language what material relations of implication and incompatibility articulate the conceptual contents of logically atomic expressions (and, as a bonus, to express the relations of implication and incompatibility that articulate the contents of the newly introduced logical expressions as well). There is no good reason to restrict the expressive ambitions with which we introduce logical vocabulary to making explicit the rare material relations of implication and incompatibility that are monotonic. Comfort with such impoverished ambition is a historical artifact of the contingent origins of modern logic in logicist and formalist programs aimed at codifying specifically *mathematical* reasoning. It is to be explained by appeal to historical causes, not good philosophical reasons.

Of course, since our tools were originally designed with this task in mind, as we have inherited them, they are best suited for the expression of monotonic rational relations. But we should not emulate the drunk who looks for his lost keys under the lamp-post rather than where he actually dropped them, just because the light is better there. We should look to shine light where we need it most.

Notice that reasons *against* a claim are as defeasible in principle as reasons *for* a claim. Material incompatibility relations are no more monotonic in general than material implication relations. Claims that are incompatible in the presence of one set of auxiliary hypotheses can in some cases be reconciled by suitable additions of collateral premises. Cases with this shape are not hard to find in the history of science.

What about Cut, the principle of cumulative *transitivity*? It is expressed in Tarski's algebraic metalanguage for consequence relations by the requirement that the consequences of the consequences of a premise-set are just the consequences of that premise-set, and by Gentzen as the principle that adding to the explicit premises of a premise-set something that is already part of its implicit content does not add to what is implied by that premise-set.

Thought of this way, Cut is the dual of what is usually thought of as the weakest acceptable structural principle that must be required if full monotonicity is not.[51] "Cautious monotonicity" is the structural requirement that adding to the explicit content of a premise-set sentences that are already part of its implicit content not defeat any implications of that premise-set. (Even though there might be *some* additional premises that *would* infirm the implication, sentences that are *already implied* by the premise-set are not among them.)

We can think generally about the structural consequences of the process of *explicitation* of content, in the sense of making what is *implicitly* contained in (or excluded by) a premise-set, in the sense of being implied by it, *explicit* as part of the explicit premises. Cut says that explicitation never *adds* implicit content. Cautious monotonicity says that explicitation never *subtracts* implicit content. Together they require that ***explicitation* is *inconsequential***. Moving a sentence from the right-hand side of the implication-turnstile to the left-hand side does not change the consequences of the premise-set. It has no effect whatever on the implicit content, on what is implied. (Explicitation can also involve making explicit what is implicitly *excluded* by a premise-set.)

---

[51] On holding onto both Cut and Cautious Monotonicity, see Gabbay, D. M., 1985, "Theoretical foundations for nonmonotonic reasoning in expert systems", in K. Apt (ed.), *Logics and Models of Concurrent Systems*, Berlin and New York: Springer Verlag, pp. 439–459. Gabbay agrees with the criteria of adequacy laid down by the influential KLM approach of Kraus, Lehman, and Magidor: Kraus, Sarit, Lehmann, Daniel, & Magidor, Menachem, 1990. Nonmonotonic Reasoning, Preferential Models and Cumulative Logics. *Artifical Intelligence*, 44: 167–207.

Explicitation in this sense is not at all a *psychological* matter. And it is not even yet a strictly *logical* notion. For even *before* logical vocabulary has been introduced, we can make sense of explicitation in terms of the structure of *material* consequence relations. Noting the effects on implicit content of adding as an explicit premise sentences that were already implied is already a process available for investigation at the semantic level of the *prelogic*.

It might well be sensible to require the inconsequentiality of explicitation as a structural constraint on *logical* consequence relations. But just as for the logical expressivist there is no good reason to restrict the rational relations of implication and incompatibility we seek to express with logical vocabulary to monotonic ones, there is no good reason to restrict our expressive ambitions to consequence relations for which explicitation is inconsequential. On the contrary, there is every reason to want to use the expressive tools of logical vocabulary to investigate cases where explicitation *does* make a difference to what is implied.

One such case of general interest is where the explicit contents of a premise-set are the records in a *database*, whose implicit contents consist of whatever consequences can be extracted from those records by applying an *inference engine* to them. (The fact that the "sentences" in the database whose material consequences are extracted by the inference engine are construed to begin with as *logically* atomic does not preclude the records having the "internal" structure of the arbitrarily complex datatypes manipulated by any object-oriented programming language.) It is by no means obvious that one is obliged to treat the results of applying the inference-engine as having exactly the same epistemic status as actual entries in the database. A related case is where the elements of the premise-sets consist of experimental *data*, perhaps measurements, or observations, whose implicit content consists of the consequences that can be extracted from them by applying a *theory*. In such a case, explicitation is far from inconsequential. On the contrary, when the CERN supercollider produces observational measurements that confirm what hitherto had been purely theoretical predictions extracted from previous data, the transformation of rational status from *mere* prediction *implicit* in prior data to actual empirical observation is an event of the first significance—no less important than the observation of something incompatible with the predictions extracted by theory from prior data. This is the very nature of empirical *confirmation* of theories. And it often happens that confirming *some* conclusions extracted by theory from the data infirms *other* conclusions that one otherwise would have drawn.

Imposing Cut and Cautious Monotonicity as global structural constraints on material consequence relations amounts to equating the epistemic status of premises and conclusions. But in many cases, we want to acknowledge a distinction, assigning a lesser status to the products of risky, defeasible inference. In an ideal case, perhaps this distinction shrinks to nothing. But we also want to be able to reason in situations where it is important to keep track of the difference in status between what we take ourselves to know and the shakier products of our theoretical reasoning from those premises. We shouldn't build into our global structural conditions on admissible material relations of implication and incompatibility assumptions that preclude us from introducing logical vocabulary to let us talk about those rational relations, so important for confirmation in empirical science.

The methodological advice not unduly to limit the structure of rational relations to which the expressive ambitions of our logics extend applies particularly forcefully to the case of incompatibility relations. The structural constraint the classical tradition for which Gentzen and Tarski speak imposes on incompatibility relations is explosion: the requirement that from incompatible premises anything and everything follows. This structural constraint corresponds to nothing whatsoever in ordinary reasoning practices, not even as institutionally codified in legal or scientific argumentative practices. It is a pure artifact of classical logical machinery, the opportune but misleading translation of the two-valued conditional into a constraint on implication and incompatibility that reflects no corresponding feature of the practices that apparatus—according to the logical expressivist—has the job of helping us to talk about. It is for that reason a perennial embarrassment to teachers of introductory logic, who are forced on this topic to adopt the low invocations of authority, pressure tactics, and rhetorical devices otherwise associated with commercial hucksters, con men, televangelists, and all the other sophists from whom since Plato we have hoped to distinguish those who are sensitive to the normative force of the better reason, whose best practices, we have since Aristotle hoped to codify with the help of logical vocabulary and its rules. In the real world, we are often obliged to reason from sets of premises that are explicitly or implicitly incompatible. [An extreme case is the legal practice of "pleading in the alternative." My defense is first, that I never borrowed the lawnmower, second, that it was broken when you lent it to me, and third that it was in perfect condition when I returned it. You have to show that *none* of these things is true. In pleading this way I am not confessing to having assassinated Kennedy. Examples from high scientific theory are not far to seek.] We should not impose structural conditions in our *prelogic* that preclude us from *logically* expressing material relations of incompatibility that characterize our actual reasoning.

III. The Expressive Role of Basic Logical Vocabulary.

The basic claim of logical expressivism in the philosophy of logic is that the expressive role characteristic of *logical* vocabulary is to make explicit, in the object-language, relations of implication and incompatibility, including the material, prelogical ones that, according to semantic inferentialism, articulate the conceptual contents expressed by nonlogical vocabulary, paradigmatically ordinary empirical descriptive vocabulary. The paradigms of logical vocabulary are the *conditional*, which codifies relations of implication that normatively structure giving reasons *for* claims, and *negation*, which codifies relations of incompatibility that normatively structure giving reasons *against* claims.

To say that a premise-set implies a conclusion, we can write in the metalanguage: "$\Gamma|\sim A$". To say that a premise-set is incompatible with a conclusion, we can write in the metalanguage "$\Gamma,A|\sim\bot$".

To perform its defining expressive task of codifying implication relations in the object language, conditionals need to satisfy the

Ramsey Condition: $\mathbf{\Gamma|\sim A\rightarrow B \text{ iff } \Gamma,A|\sim B.}$

That is, a premise-set implies a conditional just in case the result of adding the antecedent to that premise-set implies the consequent. A conditional that satisfies this equivalence can be called a "Ramsey-test conditional," since Frank Ramsey first proposed thinking of conditionals this way.

To perform its expressive task of codifying incompatibility relations in the object language, negation needs to satisfy the

Minimal Negation Condition: **$\Gamma |\!\sim \neg A$ iff $\Gamma, A |\!\sim \bot$.**

That is, a premise-set implies not-A just in case A is incompatible with that premise-set. (It follows that ¬A is the minimal incompatible of A, in the sense of being implied by everything that is incompatible with A.)

We should aspire to expressive logics built onto material incompatibility relations that are nonmonotonic as well as material implication relations that are nonmonotonic. That means that just as an implication $\Gamma |\!\sim A$ can be defeated by adding premises to Γ, so can an incompatibility. $\Gamma, A |\!\sim \bot$ can also be defeated, the incompatibility "cured", by adding some additional auxiliary hypotheses to Γ. And while, given the role negation plays in codifying incompatibilities, an incompatible set, $\Gamma \cup \{A\}$, where $\Gamma, A |\!\sim \bot$, will imply the negations of all the premises that are its explicit members, it need not therefore imply *everything*. In substructural expressive logics built on Gentzen's multisuccedent system, the condition that emerges naturally is not *ex falso quodlibet*, the classical principle of explosion, but what Ulf Hlobil calls "*ex fixo falso quodlibet*" (EFF). This is the principle that if Γ is not only materially incoherent (in the sense of explicitly containing incompatible premises) but *persistently* so, that is *incurably*, *indefeasibly* incoherent, in that all of its supersets are also incoherent, *then* it implies everything. In a monotonic setting, this is equivalent to the usual explosion principle. In nonmonotonic settings, the two conditions come apart. One conclusion that might be drawn from expressive logics is that what mattered all along was always *ex fixo falso quodlibet*—classical logic just didn’t have the expressive resources to distinguish this from explosion of all incoherent sets.

The basic idea of expressivist logic is to start with a language consisting of nonlogical (logically atomic) sentences, structured by relations of material implication and incompability. In the most general case, we think of those relations as satisfying the structural principles *only* of extended reflexivity—not monotonicity, not cautious monotonicity, and not even transitivity in the form of Cut. We then want to introduce logical vocabulary on top of such a language. This means extending the language to include arbitrarily logically complex sentences formed from the logically atomic sentences by repeatedly applying conditionals and negations, and then extending the underlying material consequence and incompatibility relations to that logically extended language in such a way that the Ramsey condition and the Minimal Negation Condition both hold. (In fact, we’ll throw in conjunction and disjunction as well.)

A basic constraint on such a construction is set out by a simple argument due to Ulf Hlobil.[52] He realized that in the context of Contexted Reflexivity and a Ramsey conditional, Cut entails Monotonicity. For if we start with some arbitrary implication Γ|~A, we can derive Γ,B|~A for arbitrary B—that is, we can show that arbitrary additions to the premise-set, arbitrary weakenings of the implication, preserves those implications. And that is just monotonicity. For we can argue:

$$\underline{\Gamma|\!\sim A} \qquad\qquad \text{Assumption}$$

---

[52] Hlobil, U. (2016), “A Nonmonotonic Sequent Calculus for Inferentialist Expressivists.” In Pavel Arazim and Michal Dančák (eds.) The *Logica* Yearbook 2015, pp. 87-105, College Publications: London.

$\underline{\Gamma, A, B \mid\!\sim A}$	Contexted Reflexivity
$\underline{\Gamma, A \mid\!\sim B \rightarrow A}$	Ramsey Condition Right-to-Left
$\underline{\Gamma \mid\!\sim B \rightarrow A}$	Cut, Cutting A using Assumption
$\Gamma, B \mid\!\sim A$	Ramsey Condition Left-to-Right.

Since we want to explore adding Ramsey conditionals to codify material implication relations that are reflexive but do not satisfy Cut—so that prelogical explicitation is not treated as always inconsequential, we will sacrifice Cut in the logical extension.

It is a minimal condition of logical vocabulary playing its defining expressive role that introducing it extend the underlying material consequence and incompatibility relations *conservatively*. (Belnap motivates this constraint independently, based on considerations raised by Prior's toxic "tonk" connective. The logical expressivist has independent reasons to insist on conservativeness: only vocabulary that conservatively extends the material relations of consequence and incompatibility on which it is based can count as *expressing* such relations explicitly.) So there should be no implications or incompatibilities involving only *old* (nonlogical) vocabulary that hold or fail to hold in the structure logically extended to include *new*, logical vocabulary, that do not hold or fail to hold already in the material base structure. Since that material base structure is in general nonmonotonic and intransitive, satisfying only contexted reflexivity, so must be the global relations of consequence and incompatibility that result from extending them by adding logical vocabulary.

IV. Basic Expressivist Logics

We now know how to do that in the context of Gentzen-style substructural proof theory. I will be summarizing technical work by recent Pitt Ph.D. Ulf Hlobil, now at Concordia University (on single-succedent systems) and current Pitt Ph.D. student Dan Kaplan (on multi-succedent systems).

We produce substructural logics codifying consequence and incompatibility relations that are not globally monotonic or transitive by modifying Gentzen's systems in three sequential stages. Gentzen's derivations all begin with what he called "initial sequents," in effect, axioms, (which will be the leaves of all *logical* proof trees) that are instances of immediate or simple reflexivity. That is, they are all of the form A|~A. We impose instead a structural rule that adds all sequents that are instances of *contexted* reflexivity—that is, (in the multisuccedent case) all sequents of the form Γ,A|~A,Θ. Making this change does not really change Gentzen's system LK of classical logic at all. For he can derive the contexted version from immediate Reflexivity by applying Monotonicity, that is Weakening (his "Thinning"). So, as others have remarked, Gentzen does not need the stronger principle of unrestricted monotonicity in order to get the full system LK of classical logic. He can make do just with the *very* restricted monotonicity principle of Contexted Reflexivity, which allows arbitrary weakening *only* of sequents that are instances of reflexivity, that is, which have some sentence that already appears on both sides of the sequent one is weakening. Since all Gentzen's initial sequents are instances of immediate reflexivity, being able to weaken them turns out to be equivalent to being able to weaken *all* logically derivable sequents. (The weakenings can be "permuted up" the proof trees past applications of connective rules in very much the same way Gentzen appeals to in proving his Cut-Elimination Hauptsatz.) Substituting the stronger version of Reflexivity for Gentzen's version accordingly allows dropping the structural requirement of Monotonicity. Contexted Reflexivity arises most naturally in Tarski's algebraic-topological way of thinking about consequence relations, as the principle that each premise-set is contained in its consequence set: $\Gamma \subseteq \text{con}(\Gamma)$.

We also do not impose Cut as a global structural constraint. But Gentzen's Cut-Elimination theorem will still be provable for all proof-trees whose leaves are instances of (now, contexted) Reflexivity. So the purely logical part of the system will still satisfy Cut.

The next step in modifying Gentzen's systems is to add axioms in the form of initial sequents relating logically atomic sentences that codify the initial base of *material* implications (and incompatibilities). Whenever some premise-set of atomic sentences Γ_0 implies an atomic sentence A, we add $\Gamma_0 |\!\sim A$ to the initial sequents that are eligible to serve as leaves of proof-trees, initiating derivations. (We require that this set of sequents, too, satisfies Contexted Reflexivity. We will be able to show that the connective rules preserve this property.) This is exactly the way Gentzen envisaged substantive axioms being added to his logical systems so that those systems could be used to codify substantive theories—for instance, when he considers the consistency of arithmetic. The crucial difference is that he required that these sequents, like those governing logically complex formulae, satisfy the structural conditions of Monotonicity and Cut—and we do not. We will introduce logical vocabulary to extend material consequence and incompatibility relations that do *not* satisfy Monotonicity, and that are *not* idempotent.

The third stage in modifying Gentzen's systems is accordingly to extend the pre-logical language to include arbitrarily logically complex sentences formed from that pre-logical vocabulary by the introduction of logical connectives. Gentzen's connective rules show how antecedent consequence and incompatibility relations governing the logically atomic base language can be systematically extended so as to govern the sentences of the logically extended language. Gentzen's own rules can be used to do this, with only minor tweaks. Like Ketonen's version of Gentzen's rules, ours are *reversible*. They are unlike the Gentzen-Ketonen rules in that we mix additive and multiplicative rules. They are all equivalent to Gentzen's own rules in the presence of a global structural rule of Monotonicity. But in nonmonotonic settings, they come apart. So, for instance, Gentzen's left rule for conjunction allows us to move from Γ,A|~C to Γ,A&B|~C. That builds in monotonicity on the left. We can't have that, since in the material base, it can happen that adding B as a further premise defeats the implication of C by Γ and A. We allow instead only the move from Γ,A,B|~C to Γ,A&B|~C. (A similar shift is needed in his right rule for disjunction: where he allows derivation of Γ|~AvB,Θ from Γ|~A,Θ, building in monotonicity on the right, we allow instead only the move from Γ|~A,B,Θ to Γ|~AvB,Θ.) I said above that from a logical expressivist point of view, for the conditional to do its defining job of codifying implication relations in the object language, it needs to satisfy the Ramsey condition. In Gentzen's setting, this amounts to the two principles:

$$\text{CP:}\quad \frac{\Gamma,A|\!\sim B}{\Gamma|\!\sim A\rightarrow B} \qquad \text{and} \qquad \text{CCP:}\quad \frac{\Gamma|\!\sim A\rightarrow B}{\Gamma,A|\!\sim B.}$$

The first is Gentzen's right-rule for the conditional. The second rule is not one of his. And it cannot be. For it is a simplifying rule. The only simplifying rule he has is Cut, and it is of the essence of his program to show that he can do without that rule: that every derivation that appeals to that single simplifying rule can be replaced by a derivation that does not appeal to it. Ketonen-style invertibility of connective rules, which makes root-first proof searches possible, though, requires not only Conditional Proof but the simplifying rule Converse Conditional Proof. And it is possible to show that this rule, too, like Cut is "admissible" in Gentzen's sense: every derivation that uses it can be replaced by a derivation that does not.

It can be shown that our versions of Gentzen's connective rules produce a *conservative extension* of any *nonmonotonic* material base consequence relation (including nonmonotonic incompatibility relations incorporated in such consequence relations) that satisfies the structural condition of Contexted Reflexivity. That is, in the absence of explicitly imposing a structural rule of Monotonicity (Weakening or Thinning) and Cut, the connective rules do not force global monotonicity. So the resulting, logically extended consequence relation is nonmonotonic. And the nonmonotonicity extends to logically complex formulae, for instance, as we have seen, in that from the fact that Γ,A|~C it does not follow that Γ,A&B|~C, so that from Γ|~A→C it does not follow that Γ|~(A&B)→C. The logical language that results permits the explicit codification using ordinary logical vocabulary of arbitrary *nonmonotonic*, insensitive material consequence relations in which prelogical explicitation is *not* inconsequential.

And yet, the system is supraclassical. All the theorems of Gentzen's system LK of classical logic can be derived in this system. For if we restrict ourselves to derivations all of whose leaves are instances of Contexted Reflexivity, that is, are of the form $\Gamma,A|\sim A,\Theta$, the result is just the theorems of classical logic. It is only if we help ourselves to initial sequents that are *not* of that form, the axioms that codify *material* relations of consequence and incompatibility, that we derive nonclassical results. **Gentzen never needed to require monotonicity, his "Thinning," as a *global* structural rule. He could just have used initial sequents that correspond to Contexted Reflexivity instead of immediate reflexivity**. That gives him all the weakening behavior he needs. Further, if we look only at sequents that are derivable *no matter what material base relation we extend*, sequents such as $\Gamma,A,A\rightarrow B|\sim B$, hence $\Gamma|\sim(A\&(A\rightarrow B))\rightarrow B$, we find that the "logic" of our system in this sense, too, is just classical logic. Perhaps not surprisingly, if, following Gentzen, we use essentially the same connective rules but restrict ourselves to *single* succedent sequents, the result is a globally nonmonotonic, intransitive *supraintuitionist* logic.[53]

I have been talking about the logical extension of nonmonotonic material *consequence* relations and not about the logical extension of nonmonotonic material *incompatibility* relations. But the latter are equally well-behaved. The multi-succedent connective rules for negation are just Gentzen's. But it is *not* the case that any materially incoherent premise-set implies every sentence. Such premise-sets imply both the sentences they explicitly contain and the negations of all those sentences. But they do not imply everything else. If a premise-set explicitly contains both A and $\neg A$ for some sentence A, *then* it implies everything. But that is because *persistently* or monotonically incoherent premise-sets explode—that is, sets that are not only incoherent themselves, but such that *every superset* of them is incoherent. This is what Ulf Hlobil calls "*ex* ***fixo*** *falso quodlibet.*" No specific stipulation to this effect needs to be made. It arises naturally out of the connective rules in the multisuccedent setting. If monotonicity held globally, *ex falso quodlibet* and *ex fixo falso quodlibet*" would be equivalent. Outside of derivations all of whose leaves are instances of contexted reflexivity, in our systems, they are not.

---

[53] We do have to add some special rules, to make up for some of the things that happen on the right in the cleaner multisuccedent system.

So in a clear sense, the *logic* is monotonic and transitive—indeed, classical or intuitionistic, depending (as with Gentzen) on whether we look at multi-succedent or single-succedent formulations—but the logically extended consequence and incompatibility relations *in general*, are not.[54] The logic of nonmonotonic consequence relations is itself monotonic. Yet it can *express*, in the logically extended object language, the nonmonotonic relations of implication and incompatibility that structure both the material, prelogical base language, and the logically compound sentences formed from them, as they behave in derivations that substantially depend on the material base relations.

Substructural expressivist logics suitable for making explicit nonmonotonic, nontransitive material consequence and incompatibility relations are accordingly not far to seek. They can easily be built by adding to Gentzen's system nonlogical axioms codifying those material relations of implication and incompatibility. It turns out that the relations of implication and incompatibility that hold in virtue of their logical form alone are still monotonic and transitive, even though the full consequence and implication relations codified by the logical connectives is not. So if you want Cut and Weakening, you can still have them—for purely *logical* consequence. Remember that from the point of view of logical expressivism, the point of introducing logical vocabulary is not what you can *prove* with it (what implications and incompatibilities hold in virtue of their logical form alone) but what you can *say* with it. Expressivist logics let us *say* a lot more than is said by their logical theorems.

---

[54] When I talk about "the logic" here this can mean *either* the theorems derivable just from instances of Contexted Reflexivity (following Gentzen) or what is implied by every premise-set for every material base relation of implication and incompatibility that satisfies Contexted Reflexivity.

V. Codifying *Local* Regions of Structure: Monotonicity as a Modality

The master-idea of logical expressivism is that logical vocabulary and the concepts such vocabulary expresses are distinguished by playing a characteristic expressive role. They let us talk, in a logically extended object language, about the *material* relations of implication and incompatibility—what is a reason for and against what—that already articulate the conceptual contents of *non*logical vocabulary, as well as the *logical* relations of implication and incompatibility built on top of those material relations. Expressivist logics are motivated by the idea that we unduly restrict the expressive power of our logics if we assume that the global structural principles that have traditionally been taken to govern purely *logical* relations of consequence and inconsistency must be taken also to govern the underlying *material* consequence and incompatibility relations. So we don't presuppose Procrustean *global* structural requirements on the material relations of consequence and incompatibility we want to codify logically. Instead, we relax those global structures and introduce vocabulary that will let us *say explicitly*, in the logically extended object language, *that* they hold *locally*, wherever in fact they still do.

Material consequence relations, I have claimed, are not in general monotonic. But they are not always and everywhere *non*monotonic. *Some* material implications are *persistent*, in that they continue to hold upon arbitrary additions to their premises. It follows from the fact that the regular planar polygon has more than three sides that its angles add up to more than 180º, no matter what additional premises we throw in. The mistake of the tradition was not to think that *there are* material implications like this, but to think that *all* material implications *must* be like this. Logical expressivists want to introduce logical vocabulary that explicitly marks the difference between those implications and incompatibilities that are persistent under the addition of arbitrary auxiliary hypotheses or collateral commitments, and those that are not. Such vocabulary lets us draw explicit boundaries around the islands of monotonicity to be found surrounded by the sea of nonmonotonic material consequences and incompatibilities.

From a Gentzenian perspective, expressivist logics work out a different way of conceiving the relations between *structure* and *connective rules*. Connectives are introduced to express local structures. The paradigm is the conditional, which codifies the implication turnstile, by satisfying the Ramsey condition in the form of CP and CCP. Conjunction codifies the comma on the left of the turnstile, and disjunction codifies the comma on the right of the turnstile (in multi-succedent systems). (Note that in our nonmonotonic setting, this requires multiplicative rather than additive rules for conjunction on the left and disjunction on the right.[55]) Negation codifies incompatibility (in Gentzen's multisuccedent systems elegantly captured in the *relation* between commas on the left and commas on the right). Our expressivist logics show how, in addition to the structures already captured by traditional connectives, further connectives can be introduced to mark local regions of the consequence relation where structure such as monotonicity and transitivity hold. I’ll try to give some idea of how this works by sketching what is for us the paradigm case: monotonicity.

---

[55] $\dfrac{\Gamma,A,B|\!\sim\Theta}{\Gamma,A\&B|\!\sim\Theta}$ and $\dfrac{\Gamma|\!\sim A,B,\Theta}{\Gamma|\!\sim A\vee B,\Theta}$ rather than $\dfrac{\Gamma,A|\!\sim\Theta}{\Gamma,A\&B|\!\sim\Theta}$ $\dfrac{\Gamma,B|\!\sim\Theta}{\Gamma,A\&B|\!\sim\Theta}$ and $\dfrac{\Gamma|\!\sim A,\Theta}{\Gamma|\!\sim A\vee B,\Theta}$ $\dfrac{\Gamma|\!\sim B,\Theta}{\Gamma|\!\sim A\vee B,\Theta}$.

The first idea is to extend the expressive power of our proof-theoretic metalanguage, so as to be able to distinguish persistent implications. In addition to the generally nonmonotonic snake turnstile “$|\!\sim$”, we can introduce a variant with an upward arrow, “$|\!\sim^{\uparrow}$” to mark persistent implications, that is, those that hold monotonically. To do this is to add *quantificational* expressive power to our proof-theoretic metalanguage. $\Gamma|\!\sim^{\uparrow}A$ says that not only does Γ imply A, but so does every superset of Γ: $\Gamma|\!\sim^{\uparrow}A$ iff $\forall X\subseteq L[\Gamma,X|\!\sim A]$.

All the connective rules can then be stipulated to have two forms: one for each turnstile. So we can write the right-rule (CP) for our Ramsey-test conditional showing the persistence arrow as optional, as:

$$\frac{\Gamma,A|\!\sim^{(\uparrow)}B}{\Gamma|\!\sim^{(\uparrow)}A\rightarrow B.}$$

If the there is no upward arrow on the top turnstile, then there is none on the bottom either. But if there is a persistence-indicating upward arrow on the premise-sequent, then there is one also on the conclusion sequent. If Γ together with A *persistently* implies B—no matter what further premises we adjoin to them—then Γ *persistently* implies the conditional—no matter what further premises we adjoin to it. That follows from the original rule, together with the definition of persistence.

From this more structurally relaxed, nonmonotonic vantage point, traditional monotonic logic looks just the way it would if there were a notationally suppressed upward arrow on all of its turnstiles.

Incompatibility (and so logical inconsistency) also looks different in this setting. We now can distinguish materially incoherent premise-sets, where $\Gamma|\sim\bot$, from *persistently* incoherent premise-sets. These are premise-sets that are not only incoherent, but whose incoherence cannot be cured by the addition of further premises. And we can restrict explosion to those *persistently* incoherent sets. If $\Gamma|\sim\bot$, then for any $A\in\Gamma$, $\Gamma|\sim A$ and $\Gamma|\sim\neg A$. But it need *not* follow that for arbitrary B, $\Gamma|\sim B$. That follows only if $\Gamma|\sim^{\uparrow}\bot$. In the single-succedent case, we stipulate this: not *ex falso quodlibet* but *ex* ***fixo*** *falso quodlibet*: ExFF. In the multi-succedent case, we do not need this stipulation. It falls out of the standard Gentzen treatment of negation. Here we want to say that what was always right about the idea that everything follows from a contradiction (and in our systems, if $A\in\Gamma$ and $\neg A\in\Gamma$, then Γ is *persistently* incoherent, and *does* imply everything) is that *persistently* incoherent premise-sets imply everything. It's just that in rigidly monotonic systems, *all* incoherence is treated as persistent.

Once the dual-turnstile apparatus is in place in the metalanguage, we can introduce a modal operator in the object language to let us say there *that* an implication holds persistently. The basic idea is to introduce a monotonicity-box that says that $\Gamma|\sim\Box A$ iff $\Gamma|\sim^{\uparrow}A$, that is, if and only if $\forall X\subseteq L[\Gamma,X|\sim A]$. To say that Γ implies $\Box A$ is just to say that Γ *persistently* (that is, monotonically) implies A. The monotonicity box is clearly a strong modality, in that if Γ implies □A, then it implies A. And it is an S4 modality, in that if Γ implies □A, then it implies □□A.

From the point of view of a globally nonmonotonic implication relation in which local pockets of monotonicity are marked in the object language by implication of modally qualified claims, the assumption of global monotonicity appears as what happens when one looks only at the monotonicity-*necessitations* of claims, ignoring anything not of the form □A.

In fact, we can do a lot better than what I have indicated so far. The expressivist idea is that the point of introducing logical vocabulary is to provide expressive resources that let one make explicit crucial local structural features of relations of implication and incompatibility—in the first instance, *material* relations of implication and incompatibility, and only as a sort of bonus the *logical* relations of implication and incompatibility that are built on top of them. From this point of view, what matters most is local persistence of some *material* implications. For it is these regions of local monotonicity in the material base relations of consequence and incompatibility that we want to be able to capture with a monotonicity-modal operator. Happily, it turns out that all we really need is an upward-arrow turnstile marking implications that can be weakened by the addition of arbitrary sets of logically *atomic* sentences. Our versions of Gentzen's connective rules then guarantee that arbitrary weakenings by sets of logically complex formulae will be possible when and only when arbitrary weakening by sets of atoms is possible according to the underlying material base consequence relation.

In addition to implications whose persistence is underwritten by peculiarities of the underlying material consequence relation, there are implications of sentences prefaced by the monotonicity box that reflect logical relations induced by the connective definitions. Sentences like these—for instance, $\Box(A \rightarrow A)$—do not depend on vagaries of the material implication relations.

A further innovation, pioneered by Ulf Hlobil for supra-intuitionistic single-succedent systems and by Dan Kaplan for supra-classical multiple-succedent systems, is the introduction of a much more powerful way of marking quantificational facts about sequents in the proof-theoretic metalanguage. (For simplicity, I'll continue to use the single-succedent case.) Instead of introducing a simple upward arrow, as I have appealed to in my sketch, we introduce an upward arrow subscripted with a set of sets. $\Gamma|\sim^{\uparrow X}A$ is defined as holding just in case for every set of sentences $X_i \in X$, $\Gamma, X_i|\sim A$. (In fact it suffices here, too, to restrict the values of X to sets of sets of logical *atoms* in the nonlogical material base language, but I put that complication aside here.) Then the set X specifies a set of sets of sentences that one can weaken Γ with, while preserving the implication of A. That is, it marks a *range of subjunctive robustness* of the implication $\Gamma|\sim A$. These are sets of sentences that can be added to Γ as collateral premises or auxiliary hypotheses without defeating the implication of A.

The underlying thought is that the most important information about a material implication is not whether or not it is monotonic—though that is something we indeed might want to know. It is rather under what circumstances it is robust and under what collateral circumstances it would be defeated. All implications are robust under *some* weakenings, and most are *not* robust under *all* weakenings. The space of material implications that articulates the contents of the nonlogical concepts those implications essentially depend upon has an intricate localized structure of subjunctive robustness and defeasibility. That is the structure we want our logical expressive tools to help us characterize. It is obscured by commitment to global structural monotonicity—however appropriate such a commitment might be for purely *logical* relations of implication and incompatibility.

Here, too, our variants of Gentzen's connective definitions, as well as those for the monotonicity box, are so contrived as to ensure that it suffices to look at ranges of subjunctive robustness of implications that are restricted to the logical atoms governed by *material* relations of consequence and incompatibility. The more fine-grained control over ranges of subjunctive robustness offered by the explicitly quantified upward arrow apparatus is governed by a couple of structural principles. To indicate their flavor: one lets us combine sets of sets under which a particular implication is robust:

$$\frac{\Gamma|\!\sim^{\uparrow X}A \qquad \Gamma|\!\sim^{\uparrow Y}A}{\Gamma|\!\sim^{\uparrow X\cup Y}A} \qquad \text{Union}$$

If the implication of A by Γ is robust under weakening by all the sets in X and it is robust under weakening by all the sets in Y, then it is robust under weakening by all the sets in $X\cup Y$. The very same connective rules stated with ordinary turnstiles go through as well with these quantified upward arrows with the same subjunctive-robustness subscript, and so propagate down proof trees.

The result of the addition of this apparatus is extensions of material consequence and incompatibility relations to a language including logically complex sentences, including those formed using the monotonicity modal box, that is well-defined and conservative of the material base relations. It follows that if the base relations are nonmonotonic and do not satisfy any version of Cut, then neither will the extended ones. The only structural principle we do impose on the base consequence relation, Contexted Reflexivity, is preserved. We do not impose the simplifying rule of Converse Conditional Proof (CCP)

$$\frac{\Gamma|\!\sim A \rightarrow B}{\Gamma, A|\!\sim B}$$

as a rule, but can prove it admissible, that is, as holding as a consequence of the connective rules for the conditional we do impose. The system is supraintuitionistic, in the single-succedent case, and supraclassical, in the multisuccedent case. If we restrict ourselves to elaborating material base consequence relations that consist entirely of instances of contexted reflexivity, that is of sequents of the form $\Gamma_0,p|\!\sim p$ for atomic sentences, then the logics over the extended languages are simply intuitionism and classical logic, respectively. These are obviously monotonic (so the monotonicity box is otiose), and Cut is, as usual, provably admissible.

VI. Conclusion

Construed narrowly, logical expressivism is a response to the demarcation question in the philosophy of logic. It suggests that we think of logical vocabulary and the concepts such vocabulary expresses as distinguished by playing a particular expressive role. The task distinctive of logical vocabulary as such is to make explicit relations of consequence and incompatibility—to allow us to *say* what claims follows from other claims, and what claims rule out others. Construed more broadly, logical expressivism invites us not to think about logic as having an autonomous subject matter—not *logical truth*, nor even *logical consequence*. Logic does not supply a canon of right reasoning, nor a standard of rationality. It takes its place in the context of an already up-and-running rational enterprise of making claims and giving reasons for and against claims. Logic provides a distinctive *organ of self-consciousness* for such a rational practice. It provides expressive tools for talking and thinking about the relations of implication and incompatibility that structure the giving of reasons for and against claims.

We should want those tools to be as broadly applicable as possible. The rational relations of material consequence that articulate the contents of nonlogical concepts are not in general monotonic. Good inferences can be infirmed be adding new information. Indeed, offering finitely statable reasons typically *requires* that the implications we invoke be defeasible. Logic should not ignore this fact, nor even aim to rectify it. Logic should aim rather to codify even nonmonotonic, intransitive reasoning.

What I have here called "expressive logics" do that. The tweaks required to the proof-theoretic apparatus Gentzen bequeathed us for it to be capable of codifying nonmonotonic, even intransitive, reasoning are remarkably small. That fact tends to confirm the expressivist's philosophical claims about what the *point* of logic has been all along. Expressive logics move beyond traditional logic not only in being built on antecedent relations of material consequence and incompatibility and in refusing to impose all but the most minimal global structural restrictions on those relations.[56] They also introduce logical vocabulary that lets one express, in the logically extended language with its logically extended relations of consequence and incompatibility, *local* regions where structural conditions *do* hold. The paradigm is the introduction of a modal operator to mark the special class of implications that *can* be arbitrarily weakened with further collateral premises (which turns out to include all those that hold in virtue of the meanings of the logical connectives alone). The benefits of treating monotonicity as a modality are many, and the costs are few. Treating logic as built on and explicating (elaborated from and explicative of) material relations of consequence and incompatibility offers another option besides substructural logics, when relaxing global structural constraints. One can introduce logical vocabulary to codify fine-grained *local* structures. These *monotonicity-modal* expressivist logics implement technically the central methodological principle of expressivist logics: don't presuppose Procrustean *global* structural requirements on the material relations of consequence and incompatibility one seeks to codify logically. Instead, relax those global structures and introduce vocabulary that will let one *say explicitly*, in the logically extended object language, *that* they hold *locally*, wherever in fact they still do.

End

---

[56] Of course not everyone—relevantists, for example—will agree that contexted reflexivity *is* minimal structure. So it should be admitted that this is a contentious description.