Commitments and Concepts

Riyadh-MANA Lectures—Winter 2022

Robert Brandom

Commitments and Concepts

What is Philosophy?[1]

Robert Brandom

We might to begin with acknowledge a distinction between things that have *natures* and things that have *histories*. Physical things such as electrons and aromatic compounds would be paradigmatic of the first class, while cultural formations such as English Romantic poetry and Ponzi schemes would be paradigmatic of the second. Applied to the case at hand, this distinction would surely place philosophy on the side of things that have histories. But now we might ask: Does philosophy differ in this respect from physics, chemistry, or biology? Physical, chemical, and biological *things* have natures rather than histories, but what about the *disciplines* that define and study them? Should physics itself be thought of as something that has a nature, or as something that has a history? Concluding the latter is giving a certain kind of pride of place to the historical. For it is in effect treating the *distinction* between things that have natures and things that have histories, between things studied by the *Naturwissenschaften* and things studied by the *Geisteswissenschaften*, as itself a cultural formation: the sort of thing that itself has a history rather than a nature. And from here it is a short step (though not, to be sure, an obligatory one) to the thought that natures themselves are the sort of thing that have a history; certainly the *concepts* electron and aromatic compound are that sort of thing. At this point the door is opened to a thorough-going historicism. It is often thought that this is the point to which Hegel—one of my particular heroes—brought us. I think that thought is correct, as far as it goes, but that we go very wrong if we think that that is where Hegel left us.

To say that philosophy is, at least to begin with, to be understood as the sort of thing that has a history rather than a nature is to foreground the way in which what deserves to be counted as distinctively *philosophical* activity answers to what has actually been done by those we recognize as precedential, tradition-transforming philosophers. One of Hegel's deepest and most important insights, I think, is indeed that the determinate contentfulness of any universal—in this

[1] Lecture material drawn from "Reason, Expression, and the Philosophical Enterprise" in *What Is Philosophy?*, C.P. Ragland and Sarah Heidt (eds.), Yale University Press, 2001, pp. 74-95.

case, the concept of philosophy—can only be understood in terms of the process by which it incorporates the contingencies of the particulars to which it has actually been applied. But he goes on from there to insist that it is in each case the responsibility of those of us who are heirs to such a conceptual tradition to see to it that is a *rational* tradition: that the distinction it embodies and enforces between correct and incorrect applications of a concept can be *justified*, that applying it in one case and withholding application in another is something for which *reasons* can be given. It is only insofar as we can do that that we are entitled to understand what we are doing as applying *concepts*. We fulfill that obligation by rationally reconstructing the tradition, finding a coherent, cumulative trajectory through it that reveals it as expressively progressive—as the gradual unfolding into greater explicitness of commitments that can be seen retrospectively as always already having been implicit in it. That is, it is our job to rewrite the history so as to discover in it the revelation of what then retrospectively appears as an antecedent nature. Hegel balances the insight that even natures have histories by seeing rationality itself as imposing the obligation to construe histories as revelatory of natures.

The aim is to pick out a sequence of precedential instances or applications of a concept that amount to the delineation of a *content* for the concept, much as a judge at common law is obliged to do. *Making* the tradition rational, is not independent of the labor of concretely *taking* it to be so. It is a criterion of adequacy of each such Whiggish rewriting of our disciplinary history that it create and display continuity and progress by its systematic inclusions and exclusions. The discontinuities that correspond to shifts of topic, the forgetting of lessons, and the degeneration of research programs are invisible from within each such telling; but those differences live on in the spaces between the tellings. Each generation redefines its subject by offering a new retrospective reading of its characteristic concerns and hard-won lessons. But also, at any one time there will be diverse interpretations, complete with rival canons, competing designations of heroes, and accounts of their heroic feats. Making canons and baking traditions out of the rich ingredients bequeathed us by our discursive predecessors is a game that all can play.

In this talk, I am going to sketch one such perspective on what philosophers do— by discerning a nature as revealed by the history.

Ours is a broadly *cognitive* enterprise—I say '*broadly* cognitive' to indicate that I mean that philosophers aim at a kind of *understanding*, not, more narrowly, at a kind of *knowledge*. To specify the distinctive sort of understanding that is the characteristic goal of philosophers' writing is to say what distinguishes that enterprise from that of other sorts of constructive seekers of understanding, such as novelists and scientific theorists. I want to do so by focusing not on the peculiar genre of nonfiction creative writing by which philosophical understanding is typically conveyed (though I think that subject is worthy of consideration), but rather on what is distinctive about the understanding itself: both its particular *topic*, and its characteristic *goal*. Philosophy is a self-reflexive enterprise: understanding is not only the *goal* of philosophical inquiry, but its *topic* as well. *We* are its topic; but it is us specifically as *understanding* creatures: *discursive* beings, makers and takers of *reasons*, seekers and speakers of *truth*. Seeing philosophy as addressing the nature and conditions of our rationality is, of course, a very traditional outlook—so traditional, indeed, that it is liable to seem quaint and old-fashioned. I'll address this issue later, remarking now only that *rationalism* is one thing, and *intellectualism* another: pragmatists, too, are concerned with the practices of giving and asking for reasons.

I understand the task of philosophers to have as a central element the explication of concepts—or, put slightly more carefully, the development and application of expressive tools with which to make explicit what is implicit in the use of concepts. When I say "explication of concepts", it is hard not to hear "analysis of meanings." There are obviously affinities between my specification and that which defined the concern specifically of "analytic philosophy" in the middle years of the last century. Indeed, I intend, *inter alia*, to be saying what was right about that conception. But what I have in mind is different in various ways. *Explication*, making explicit, is not the same as *analysis*, at least as that notion was classically conceived. As I use the term, for instance, we have no more privileged access to the contents of our concepts than we do to the facts we use them to state; the concepts and the facts are two sides of one coin.

But the most important difference is that where analysis of meanings is a fundamentally *conservative* enterprise (consider the paradox of analysis), I see the point of explicating concepts rather to be opening them up to rational *criticism*. The rational enterprise, the practice of giving

and asking for reasons that lies at the heart of discursive activity, requires not only criticizing *beliefs*, as false or unwarranted, but also criticizing *concepts*. Defective concepts distort our thought and constrain us by limiting the propositions and plans we can entertain as candidates for endorsement in belief and intention. This constraint operates behind our backs, out of our sight, since it limits what we are so much as capable of being aware of. Philosophy, in developing and applying tools for the rational criticism of concepts, seeks to free us from these fetters, by bringing the distorting influences out into the light of conscious day, exposing the commitments implicit in our concepts as vulnerable to rational challenge and debate.

I

The first thing to understand about concepts is that concept is a *normative* concept. This is a lesson we owe ultimately to Kant—the great, gray mother of us all. Kant saw us above all as traffickers in concepts. In fact, in a strict sense, *all* that kantian rational creatures can do *is* to apply concepts. For that is the genus he took to comprise both *judgment* and *action*, our theoretical activity and our practical activity. One of Kant's great innovations was his view that what in the first instance distinguishes judgments and actions from the mere behavior of denizens of the realm of nature is that they are things that we are in a distinctive sense *responsible* for. They express *commitments* of ours. The norms or rules that determine what we have committed ourselves to, what we have made ourselves responsible for, by making a judgment or performing an action, Kant calls 'concepts'. Judging and acting involves undertaking commitments whose credentials are always potentially at issue. That is, the commitments embodied in judgments and actions are ones we may or may not be *entitled* to, so that the question of whether they are *correct*, whether they are commitments we *ought* to acknowledge and embrace, can always be raised. One of the forms taken by the responsibility we undertake in judging and acting is the responsibility to give reasons that justify the judgment or the action. And the rules that are the concepts we apply in judging and acting determine what would count as a reason for the judgment and the action.

Commitment, entitlement, responsibility—these are all normative notions. Kant replaces the *ontological* distinction between the physical and the mental with the *deontological* distinction between the realm of nature and the realm of freedom: the distinction between things that merely act regularly and things that are subject to distinctively normative sorts of assessment.

Thus for Kant the great philosophical questions are questions about the source and nature of normativity—of the bindingness or validity [Gültigkeit] of conceptual rules. Descartes had bequeathed to his successors a concern for *certainty*: a matter of our grip on concepts and ideas—paradigmatically, whether we have a hold on them that is clear and distinct. Kant bequeaths to his successors a concern rather for *necessity*: a matter of the grip concepts have on

us, the way they bind or oblige us. 'Necessary' [notwendig] for Kant just means "according to a rule". (That is why he is willing to speak of moral and natural necessity as species of a genus.) The important lesson he takes Hume to have taught isn't about the threat of skepticism, but about how empirical knowledge is unintelligible if we insist on merely *de*scribing how things in fact *are*, without moving beyond that to *pre*scribing how they *must* be, according to causal rules, and how empirical motivation (and so agency) is unintelligible if we stay at the level of '*is*' and eschew reference to the '*ought*'s that outrun what merely is. Looking farther back, Kant finds "the celebrated Mr. Locke" sidetracked into a mere "physiology of the understanding"—the tracing of causal antecedents of thought in place of its justificatory antecedents—through a failure to appreciate the essentially normative character of claims to knowledge. But Kant takes the whole Enlightenment to be animated by an at least implicit appreciation of this point. For mankind's coming into its intellectual and spiritual majority and maturity consists precisely in taking the sort of personal responsibility for its commitments, both doxastic and practical, insisted upon already by Descartes' meditator.

This placing of normativity at the center of philosophical concern is the reason behind another of Kant's signal innovations: the pride of place he accords to *judgment*. In a sharp break with tradition, he takes it that the smallest unit of experience, and hence of awareness, is the judgment. This is because judgments, applications of concepts, are the smallest unit for which knowers can be *responsible*. Concepts by themselves don't express commitments; they only determine what commitments would be undertaken if they were applied. (Frege will express this kantian point by saying that judgeable contents are the smallest unit to which pragmatic force—paradigmatically the assertional force that consists in the assertor undertaking a special kind of commitment—can attach. Wittgenstein will distinguish sentences from terms and predicates as the smallest expressions whose free-standing utterance can be used to make a move in a language game.) The most general features of Kant's understanding of the form of judgment also derive from its role as a unit of responsibility. The "I think" that can accompany all representations (hence being, in its formality, the emptiest of all) is the formal shadow of the transcendental unity of apperception, the locus of responsibility determining a coresponsibility class of concept-applications (including actions), what is responsible *for* its judgments. The objective correlate of this subjective aspect of the form of judgment is the "object=X" to which

the judgment is directed, the formal shadow of what the judgment makes the knower responsible *to*.

I think that philosophy is the study of us as creatures who judge and act, that is, as discursive, concept-using creatures. And I think that Kant is right to emphasize that understanding what we do in these terms is attributing to us various kinds of normative status, taking us to be subject to distinctive sorts of normative appraisal. So a central philosophical task is understanding this fundamental *normative* space in which we live, and move, and have our being. Kant's own approach to this issue, developing themes from Rousseau, is based on the thought that genuinely normative authority (constraint by norms) is distinguished from causal power (constraint by facts) in that it binds only those who *acknowledge* it as binding. (Rousseau said that freedom is obedience to a law on lays down for oneself.) Because one is subject only to that authority one subjects oneself to, the normative realm can be understood equally as the realm of *freedom*. So being constrained by norms is not only compatible with freedom—properly understood, it can be seen to be what freedom *consists in*. I don't know of a thought that is deeper, more difficult, or more important than this.

II

Kant's most basic idea, I said, is that judgment and action are things we are in a distinctive way *responsible* for. What does it mean to be responsible for them? I think the kind of responsibility in question should be understood to be a *task* responsibility: the responsibility to do something. What (else) do judging and acting oblige us to do? The commitments we undertake by applying concepts in particular circumstances—by judging and acting—are ones we may or may not be entitled to, according to the rules (norms) implicit in those concepts. Showing that we are entitled by the rules to apply the concept in a particular case is *justifying* the commitment we undertake thereby, offering *reasons* for it. That is what we are responsible for, the practical content of our conceptual commitments. In undertaking a conceptual commitment one renders oneself in principle liable to demands for reasons. The normative appraisal to which we subject ourselves in judging and acting is appraisal of our reasons. Further, offering a reason for the application of a concept is always applying another concept: making or rehearsing another judgment or undertaking or acknowledging another practical commitment (Kant's "adopting a maxim"). Conceptual commitments both serve as and stand in need of reasons. The normative realm inhabited by creatures who can judge and act is not only the realm of freedom, it is the realm of reason.

Understanding the norms for correct application that are implicit in concepts requires understanding the role those concepts play in reasoning: what (applications of concepts) count as reasons for the application of that concept, and what (applications of concepts) the application of that concept counts as a reason for. For apart from such understanding, one cannot fulfill the responsibility one undertakes by making a judgment or performing an action. So what distinguishes concept-using creatures from others is that we know our way around the *space of reasons*. Grasping or understanding a concept just is being able practically to place it in a network of inferential relations: to know what is evidence for or against its being properly applied to a particular case, and what its proper applicability to a particular case counts as evidence for or against. Our capacity to know (or believe) *that* something is the case depends on our having a certain kind of know *how*: the ability to tell what is a reason for and against what.

The cost of losing sight of this point is to assimilate genuinely conceptual activity, judging and acting, too closely to the behavior of mere animals—creatures who do not live and move and have their being in the normative realm of freedom and reason. We share with other animals (and for that matter, with bits of automatic machinery) the capacity reliably to respond differentially to various kinds of stimuli. We, like they, can be understood as classifying stimuli as being of certain kinds, insofar as we are disposed to produce different repeatable sorts of responses to those stimuli. We can respond differentially to red things by uttering the noise "That is red." A parrot could be trained to do this, as pigeons are trained to peck at a different button when shown a red figure than when shown a green one. The *empiricist* tradition is right to emphasize that our capacity to have empirical knowledge begins with and crucially depends on such reliable differential responsive dispositions. But though the story begins with this sort of classification, it does not end there. For the *rationalist* tradition is right to emphasize that our classificatory responses count as applications of concepts, and hence as so much as candidates for knowledge, only in virtue of their role in reasoning. The crucial difference between the parrot's utterance of the noise "That is red," and the (let us suppose physically indistinguishable) utterance of a human reporter is that for the latter, but not the former, the utterance has the practical significance of making a claim. Doing that is taking up a *normative stance* of a kind that can serve as a premise from which to draw conclusions. That is, it can serve as a reason for taking up other stances. And further, it is a stance that itself can stand in need of reasons, at least if challenged by the adoption of other, incompatible stances. Where the parrot is merely responsively sounding off, the human counts as applying a concept just insofar as she is understood as making a move in a game of giving and asking for reasons.

The most basic point of Sellars' rationalist critique of empiricism in his masterwork "Empiricism and the Philosophy of Mind," is that even the *non*inferentially elicited perceptual judgments that the empiricist rightly appreciates as forming the empirical basis for our knowledge can count as judgments (applications of concepts) at all only insofar as they are *inferentially* articulated. Thus the idea that there could be an autonomous language game (a game one could play though one played no other) consisting entirely of noninferentially elicited reports—whether of environing stimuli or of the present contents of one's own mind—is a

radical mistake. To apply any concepts *non*inferentially, one must be able also to apply concepts inferentially. For it is an essential feature of concepts that their applications can both serve as and stand in need of reasons. Making a report or a perceptual judgment is doing something that essentially, and not just accidentally, has the significance of making available a premise for reasoning. Learning to observe requires learning to infer. Experience and reasoning are two sides of one coin, two capacities presupposed by concept use that are in principle intelligible only in terms of their relations to each other.

To claim that what distinguishes specifically conceptual classification from classification merely by differential responsive disposition is the inferential articulation of the response—that applications of concepts are essentially what can both serve as and stand in need of reasons—is to assign the game of giving and asking for reasons a preeminent place among discursive practices. For it is to say that what makes a practice *discursive* in the first place is that it incorporates reason-giving practices. Now of course there are many things one can do with concepts besides using them to argue and to justify. And it has seemed perverse to some post-Enlightenment thinkers in any way to privilege the rational, cognitive dimension of language use. But if the tradition I have been sketching is right, the capacity to use concepts in all the other ways explored and exploited by the artists and writers whose imaginative enterprises have rightly been admired by romantic opponents of logocentrism is parasitic on the prosaic inferential practices in virtue of which we are entitled to see concepts as in play in the first place. The game of giving and asking for reasons is not just one game among others one can play with language. It is the game in virtue of the playing of which what one has qualifies as *language* (or thought) at all. I am here disagreeing with Wittgenstein, when he claims that "language has no downtown." On my view, it does, and that downtown (the region around which all the rest of discourse is arrayed as dependent suburbs), is the practices of giving and asking for reasons. This is a kind of linguistic *rationalism*. 'Rationalism' in this sense does not entail intellectualism, the doctrine that every *implicit* mastery of a propriety of practice is ultimately to be explained by appeal to a prior *explicit* grasp of a principle. It is entirely compatible with the sort of pragmatism that sees things the other way around.

III

As I am suggesting that we think of them, concepts are broadly inferential norms that implicitly govern practices of giving and asking for reasons. Dummett has suggested a useful model for thinking about the inferential articulation of conceptual contents. Generalizing from the model of meaning Gentzen introduces for sentential operators, Dummett suggests that we think of the use of any expression as involving two components: the circumstances in which it is appropriately used and the appropriate consequences of such use. Since our concern is with the application of the concepts expressed by using linguistic expressions, we can render this as the circumstances of appropriate application of the concept, and the appropriate consequences of such application—that is, what follows from the concept's being applicable.

Some of the circumstances and consequences of applicability of a concept may be inferential in nature. For instance, one of the circumstances of appropriate application of the concept red is that this concept is applicable wherever the concept scarlet is applicable. And to say that is just another way of saying that the inference from "X is scarlet," to "X is red," is a good one. And similarly, one of the consequences of the applicability of the concept red is the applicability of the concept colored. And to say that is just another way of saying that the inference from "X is red," to "X is colored," is a good one. But concepts like red also have *non*inferential circumstances of applicability, such as the visible presence of red things. And concepts such as unjust have noninferential consequences of application—that is, they can make it appropriate to *do* (or not do) something, to make another claim true, not just to *say* or judge that it is true.

Even the immediately empirical concepts of *observables*, which have noninferential *circumstances* of application and the immediately practical *evaluative* concepts, which have noninferential *consequences* of application, however, can be understood to have contents that are inferentially articulated. For all concepts incorporate an implicit commitment to the propriety of the inference from their circumstances to their consequences of application. One cannot use the concept red as including the circumstances and consequences mentioned above without

committing oneself to the correctness of the inference from "X is scarlet," to "X is colored." So we might decompose the norms that govern the use of concepts into three components: circumstances of appropriate application, appropriate consequences of application, and the propriety of an inference from the circumstances to the consequences. I would prefer to understand the inferential commitment expansively, as including the circumstances and consequences it relates, and so as comprising all three normative elements.

I suggested at the outset that we think of philosophy as charged with producing and deploying tools for the criticism of concepts. The key point here is that concepts may incorporate defective inferences. Dummett offers this suggestive example:

> A simple case would be that of a pejorative term, e.g. 'Boche'. The conditions for applying the term to someone is that he is of German nationality; the consequences of its application are that he is barbarous and more prone to cruelty than other Europeans. We should envisage the connections in both directions as sufficiently tight as to be involved in the very meaning of the word: neither could be severed without altering its meaning. Someone who rejects the word does so because he does not want to permit a transition from the grounds for applying the term to the consequences of doing so.[2]

(It is useful to focus on a French epithet from the first world war, because we are sufficiently removed from its practical effect to be able to get a theoretical grip on how it works. But the thought should go over *mutatis mutandis* for pejoratives in current circulation.) Dummett's idea is that if you do not accept as correct the inference from German nationality to an unusual disposition to barbarity and cruelty, you can only reject the word. You cannot deny that there are any Boche, for that is just denying that the circumstances of application are ever satisfied, that is, that there are any Germans. And you cannot admit that there are Boche but deny that they are disposed to barbarity and cruelty (this is the "Some of my best friends are Boche," ploy), since that is just taking back in one breath what one has asserted just before. Any use of the term commits the user to the inference that is curled up, implicitly, in it. (At Oscar Wilde's trial the prosecutor read out some passages from the Importance of Being Earnest and said "I put it to

[2] Dummett, *Frege: Philosophy of Language* [Harper and Row, New York, 1973] p. 454.

you, Mr. Wilde, that this is blasphemy. Is it? Yes or no?" Wilde replied just as he ought on the account I am urging: "Sir, 'blasphemy' is not one of my words."[3])

Although they are perhaps among the most dangerous, it is not just highly-charged words, words that couple 'descriptive' circumstances of application with 'evaluative' consequences of application that incorporate inferences of which we may need to be critical. The use of *any* expression involves commitment to the propriety of the inference from its circumstances to its consequences of application. These are almost never logically valid inferences. On the contrary, they are what Sellars called "material" inferences: inferences that articulate the content of the concept expressed. Classical disputes about the nature of personal identity, for instance, can be understood as taking the form of arguments about the propriety of such a material inference. We can agree, we may suppose, about the more or less forensic consequences of application of the concept "same person," having in mind its significance for attributions of (co-)responsibility. When we disagree about the circumstances of application that should be paired with it—for instance whether bodily or neural continuity, or the psychological continuity of memory count for more—we are really disagreeing about the correctness of the inference from the obtaining of these conditions to the ascription of responsibility. The question about what is the correct concept is a question about which inferences to endorse. I think it is helpful to think about a great number of the questions we ask about other important concepts in these same terms: as having the form of queries about what inferences from circumstances to consequences of application we ought to acknowledge as correct, and why. Think in these terms about such very abstract concepts as morally wrong, just, beautiful, true, explain, know, or prove, and again about 'thicker' ones such as unkind, cruel, elegant, justify, and understand.

The use of any of these concepts involves a material inferential commitment: commitment to the propriety of a substantial inferential move from the circumstances in which it is appropriate to apply the concept to the consequences of doing so. The concepts are substantive just because the inferences they incorporate are. Exactly this commitment becomes invisible, however, if one conceives conceptual content in terms of *truth conditions*. For the idea of truth conditions is the idea of a single set of conditions that are at once necessary and

[3] Of course, being right on this point didn't keep Wilde out of trouble, anymore than it did Salman Rushdie.

sufficient for the application of the concept. The idea of individually necessary conditions that are also jointly sufficient is the idea of a set of consequences of application that can also serve as circumstances of application. Thus the circumstances of application are understood as already including the consequences of application, so that no endorsement of a substantive inference is involved in using the concept. The concept of concepts like this is not incoherent. It is the ideal of *logical* or *formal* concepts. Thus it is a criterion of adequacy for introducing logical connectives that they be inferentially conservative: that their introduction and elimination rules be so related that they permit no new inferences involving only the old vocabulary. But it is a bad idea to take this model of the relation between circumstances and consequences of application of logical vocabulary and extend it to encompass also the substantively contentful *non*logical concepts that are the currency in which most of our cognitive and practical transactions are conducted.

It is a bad idea because of its built-in conservatism. Understanding meaning or conceptual content in terms of truth conditions—individually necessary and jointly sufficient conditions—squeezes out of the picture the substantive inferential commitment implicit in the use of any nonlogical concept. But it is precisely those inferential commitments that are subject to *criticism* in the light of substantive collateral beliefs. If one does not believe that Germans are distinctively barbarous or prone to cruelty, then one must not use the concept Boche, just *because* one does not endorse the substantive material inference it incorporates. On the other model, this diagnosis is not available. The most one can say is that one does not know how to specify truth conditions for the concept. But just what is objectionable about it and why does not appear from this theoretical perspective. Criticism of concepts is always criticism of the inferential connections. For criticizing whether all the individually sufficient conditions (circumstances) "go together", i.e. are circumstances of application of one concept, just is wondering whether they all have the same consequences of application (and similarly for wondering whether the consequences of application all "go together").

IV

When we think of conceptual contents in the way I am recommending, we can see not only how beliefs can be used to criticize concepts, but also how concepts can be used to criticize beliefs. For it is the material inferences incorporated in our concepts that we use to elaborate the antecedents and consequences of various candidates for belief—to tell what we would be committing ourselves to, what would entitle us to those commitments, what would be incompatible with them, and so on. Once it is accepted that the inferential norms implicit in our concepts are in principle as revisable in the light of evidence as particular beliefs, conceptual and empirical authority appear as two sides of one coin. Rationally justifying our concepts depends on finding out about how things are—about what actually follows from what—as is most evident in the case of massively defective concepts such as Boche.

Adjusting our beliefs in the light of the connections among them dictated by our concepts, and our concepts in the light of our evidence for the substantive beliefs presupposed by the inferences they incorporate, is the rationally reflective enterprise introduced to us by Socrates. It is what results when the rational, normative connections among claims that govern the practice of giving and asking for reasons are themselves brought into the game, as liable to demands for reasons and justification. Saying or thinking something, making it explicit, consists in applying concepts, thereby taking up a stance in the space of reasons, making a move in the game of giving and asking for reasons. The structure of that space, of that game, though, is not given in advance of our finding out how things are with what we are talking about. For what is *really* a reason for what depends on how things *actually* are. But that inferential structure itself can be the subject of claims and thoughts. It can itself be made explicit in the form of claims about what follows from what, what claims are evidence for or against what other claims, what else one would be committing oneself to by making a certain judgment or performing a certain action. So long as the commitment to the propriety of the inference from German nationality to barbarity and unusual cruelty remains merely implicit in the use of term such as ‘Boche’, it is hidden from rational scrutiny. When it is made explicit in the form of the conditional claim “Anyone who is German is barbarous and unusually prone to cruelty,” it is subject to rational

challenge and assessment; it can, for instance, be confronted with such counterexamples as Bach and Goethe.

Discursive explicitness, the application of concepts, is Kantian apperception or consciousness. Bringing into discursive explicitness the inferentially articulated conceptual norms in virtue of which we can be conscious or discursively aware of anything at all is the task of reflection, or self-consciousness. This is the expressive task distinctive of philosophy. Of course, the practitioners of special disciplines, such as membrane physiology, are concerned to unpack and criticize the inferential commitments implicit in using concepts such as lipid soluble with a given set of circumstances and consequences of application, too. It is the emphasis on the "anything at all" distinguishes philosophical reflection from the more focused reflection that goes on within such special disciplines. Earlier I pinned on Kant a view that identifies us as distinctively *rational* creatures, where that is understood as a matter of our being subject to a certain kind of *normative* assessment: we are creatures who can undertake *commitments* and *responsibilities* that are *conceptually* articulated in that their contents are articulated by what would count as *reasons* for them (as well as what other commitments and responsibilities they provide reasons for). One of philosophy's defining obligations is to supply and deploy an expressive toolbox, filled with concepts that help us make explicit various aspects of *rationality* and *normativity* in general. **The topic of philosophy is normativity in all its guises, and inference in all its forms.** And its task is an *expressive*, *explicative* one. So it is the job of practitioners of the various philosophical subfields to design and produce specialized expressive tools, and to hone and shape them with use. At the most general level, *inferential* connections are made explicit by *conditionals*, and their *normative* force is made explicit by *deontic* vocabulary. Different branches of philosophy can be distinguished by the different sorts of inference and normativity they address and explicate, the various special senses of "if…then___," or of 'ought' for which they care. Thus philosophers of science, for instance, develop and deploy conditionals codifying causal, functional, teleological, and other explanatory inferential relations, value theorists sharpen our appreciation of the significance of the differences in the endorsements expressed by prudential, legal, ethical, and aesthetic 'ought's, and so on.

V

I said at the beginning of my remarks that I thought of philosophy as defined by its history, rather than by its nature, but that, following Hegel, I think of our task as understanding it by finding or making a nature in or from its history. The gesture I have made in that direction today, though, could be also be summarized in a different kind of definition, namely in the ostensive definition: Philosophy is the kind of thing that Kant and Hegel did (one might immediately want to add Plato, Aristotle, Frege and Wittgenstein to the list, and then we are embarked on the enterprise of turning a gesture into a story, indeed, a history). So one might ask: Why not just say that, and be done with it? While, as I've indicated, I think that specification is a fine place to start, I also think there is a point to trying to be somewhat more explicit about just what sort of thing it is that one takes it Kant and Hegel (and Frege and Wittgenstein) did. Doing that is not being satisfied just with a wave at philosophy as something that has a history. It is trying rationally to reconstruct that tradition, to recast it into a form in which a constellation of ideas can be seen to be emerging, being expressed, refined, and developed.

With those giants, I see philosophy as a discipline whose distinctive concern is with a certain kind of *self-consciousness*: awareness of ourselves as specifically *discursive* (that is, concept-mongering) creatures. It's task is understanding the conditions, nature, and consequences of conceptual norms and the activities—starting with the social practices of giving and asking for reasons—that they make possible and that make them possible. As concept users, we are beings who can make explicit how things are and what we are doing—even if always only in relief against a background of implicit circumstances, conditions, skills, and practices. Among the things on which we can bring our explicitating capacities to bear are those very concept using capacities that make it possible to make anything at all explicit. Doing that, I am saying, is philosophizing.

It is easy to be misled by the homey familiarity of these sentiments, and correspondingly important to distinguish this characterization from some neighbors with which it is liable to be

confused. There is a clear affinity between this view and Kant's coronation of philosophy as "queen of the sciences." For on this account philosophy does extend its view to encompass all activity that is discursive in a broad sense—that is, all activity that presupposes a capacity for judgment and agency, sapience in general. But in this sense, philosophy is at most *a* queen of the sciences, not *the* queen. For the magisterial sweep of its purview does not serve to distinguish it from, say, psychology, sociology, history, literary or cultural criticism, or even journalism. What distinguishes it is the *expressive* nature of its concern with discursiveness in general, rather than its inclusive scope. My sketch was aimed at introducing a specific difference pertaining to philosophy, not a unique privilege with respect to such other disciplines.

Again, as I have characterized it, philosophy does not play a *foundational* role with respect to other disciplines. Its claims do not stand prior to those of the special sciences in some order of ultimate justification. Nor does philosophy sit at the other end of the process as final judge over the propriety of judgments and actions—as though the warrant of ordinary theoretical and practical applications of concepts remained somehow provisional until certified by philosophical investigation. And philosophy as I have described it likewise asserts no methodological privilege or insight that potentially collides with the actual procedures of other disciplines.

Indeed, philosophy's own proper concerns with the nature of normativity in general, and with its conceptual species in particular, so on inference and justification in general, impinge on the other disciplines in a role that equally well deserves the characterization of "handmaiden." For what we do that has been misunderstood as having foundational or methodological significance is provide and apply tools for unpacking the substantive commitments that are implicit in the concepts deployed throughout the culture, including the specialized disciplines of the high culture. Making those norms and inferences explicit in the form of claims exposes them for the first time to reasoned assessment, challenge, and defense, and so to the sort of rational emendation that is the primary process of conceptual evolution. But once the implicit presuppositions and consequences have been brought out into the daylight of explicitness, the process of assessment, emendation, and so evolution is the business of those whose concepts they are—and not something philosophers have any particular authority over or expertise

regarding. Put another way, it is the business of philosophers to figure out ways to increase semantic and discursive self-consciousness. What one does with that self-consciousness is not our business *qua* philosophers—though of course, *qua* intellectuals generally, it may well be.

Philosophy's *expressive* enterprise is grounded in its focus on us as a certain kind of thing, an expressing thing: as at once creatures and creators of conceptual norms, producers and consumers of reasons, beings distinguished by being subject to the peculiar normative force of the better reason. Its concern with us as specifically *normative* creatures sets philosophy off from the empirical disciplines, both the natural and the social sciences. It is this normative character that binds together the currents of thought epitomized in Stanley Cavell's characteristically trenchant aphorism that Kant depsychologized epistemology, Frege depsychologized logic, and Wittgenstein depsychologized psychology. We might add that Hegel depsychologized history. The depsychologizing move in question is equally a desociologizing. For it is a refocusing on the *normative bindingness* of the concepts deployed in ground-level empirical knowledge, reasoning, and thought in general. This is a move beyond the narrowly *natural* (in the sense of the describable order of causes), toward what Hegel called the 'spiritual' [geistig], that is, the *normative* order. That its concern is specifically with our *conceptual* normativity sets philosophy off from the other humanistic disciplines, from the literary as well as the plastic arts. Conceptual commitments are distinguished by their inferential articulation, by the way they can serve as reasons for one another, and by the way they stand in need of reasons, their entitlement always potentially being at issue. Now in asserting the centrality and indispensability, indeed, the criterial role, of practices of giving and asking for reasons, I am far from saying that reasoning—or even thinking—is all anyone ought to do. I am saying that philosophers' distinctive concern is with what else those reason-mongering practices make possible, and how they do, on the one hand, and with what it is that makes them possible—what sort of doings count as sayings, how believing or saying that is founded on knowing how—on the other. It is this distinctive constellation of concerns that makes philosophy the party of reasons, and philosophers the friends of the norms, the ones who bring out into the light of discursive explicitness our capacity to make things discursively explicit.

End of Lecture 1

Intentionality and Language:

A Normative, Pragmatist, Inferentialist Approach[4]

Robert B. Brandom

I. Intentionality

In this talk I present a battery of concepts, distinctions, terminology, and questions that are common currency among philosophers of mind and language who think about intentionality. Together, they define a space of possible explanatory priorities and strategies. In addition, I sketch a systematic, interlocking set of commitments regarding the relations among these concepts and distinctions, which underwrites a distinctive set of answers to some of the most important of those questions. This normative, pragmatist, inferentialist approach to intentionality and language is much more controversial. I have developed and expounded it in a number of books over the past two decades. In the present context its exposition can serve at least to illustrate how one might assemble a framework within which to think about the relations among these important issues.

The contemporary philosophical use of the medieval scholastic term "intentionality" was introduced by Franz Brentano. His student Edmund Husserl recognized it as apt to characterize a phenomenon that Immanuel Kant had put at the center of our thought about mindedness, as part of what we would now call his *semantic* transformation of René Descartes's *epistemological* turn in the philosophy of mind. This is the idea of a kind of *contentfulness* that is distinctive of at least some of our psychological states and linguistic utterances. Brentano characterized intentionality in terms of "reference to a content, a direction upon an object."[5] John Searle offers this pre-theoretical summary of the subject-matter of his book *Intentionality*:

---

[4] The original essay from which this lecture was drawn was published in N. J. Enfield, Paul Kockelman, Jack Sidnell (eds.) (2014) *The Cambridge Handbook of Linguistic Anthropology* [Cambridge University Press, 2014], as Chap. 14.

[5] Franz Brentano, "Psychology from the Empirical Standpoint," trans. D.B.Terrell, quoted on pp. 119-20 in H. Morick (ed.) *Introduction to the Philosophy of Mind: Readings from Descartes to Strawson* [Scott, Foresman; Glenview, Ill. 1970].

> ...if a state S is Intentional then there must be an answer to such questions as:
> What is S about? What is S of? What is it an S that?[6]

We can specify the content of someone's belief by saying, for instance, that she believes *that* Lampe was Kant's faithful servant. In that case, it is a belief *of* or *about* Kant's servant Lampe, *representing* him as being faithful. Brentano was impressed by the thought that while things can only stand in physical or causal relations to actually existing facts, events, and objects, intentional states can "refer to contents" that are not true (do not express actual facts) and be "directed upon objects" that do not exist.[7] I can only kick the can if it exists, but I can think about unicorns even if they do not.

We should distinguish *intentionality* in this sense from *consciousness*. These phenomena only overlap. For, on the one hand, pain is a paradigmatically conscious phenomenon. But pains are not in the sense relevant to intentionality *contentful* states or episodes. They do not have contents that could be expressed by sentential 'that' clauses. And they are not (at least not always) *about* anything. On the other hand, there is nothing incoherent about the concept of *unconscious* beliefs—which do have intentional contents specifiable both in terms of 'that' and 'of'. Attributions of belief answer to two kinds of norms of evidence, which in some cases diverge. Evidence derived from sincere avowals by the believer license the attributions of beliefs of which the believer is conscious. But beliefs, desires, and other intentional states can also be attributed on the basis of what relatively stable beliefs and desires provide premises for bits of practical reasoning that make the most sense of what the believer actually does, even in the absence of dispositions sincerely to avow the intentional states in question. Where such intentional explanations are good explanations, the attributed intentional states are unconscious.

The need to make this distinction is a manifestation of a deeper distinction between two sorts of mindedness: *sentience* and *sapience*. Sentience is awareness in the sense of being awake. Anything that can feel pain is sentient. Sapience is having intentionally contentful states such as beliefs, desires, and intentions: believing, desiring, or intending *of* the dog *that* it is sitting, will sit, or should sit. An essential element of Descartes's invention of a distinctively modern conception of the mind was his assimilation of *sensations* (for instance, pain) and

---

[6] John Searle, *Intentionality* [Cambridge University Press, 1983].

[7] Notice that it is at least not obvious that the first part of this claim is true. Reinforcing the dam might have averted a possible disaster. If so, the nonexistence of the disaster was presumably an effect caused by the reinforcement.

thoughts (for instance, that foxes are nocturnal ominivores). His predecessors had not been tempted by such an assimilation of sentience and sapience. His innovation, and the rationale for the assimilation, was an *epistemic* criterion of demarcation of the mental. Both sensations and thoughts, he took it, were *transparent* and *incorrigible* to their subject: they could not occur without the subject knowing that they occurred, and if the subject took it that they occurred, then they did. Apart from growing appreciation (beginning already with Gottfried Leibniz) of the potential explanatory significance of unconscious mental states, concerning which subjects do *not* have the sort of privileged epistemic access Descartes focused on, we have come to appreciate the importance of not prejudging issues concerning the relations between sentience and sapience. In particular, we have come to see that some of the most important issues concerning the plausibility, and even the intelligibility, of artificial intelligence as classically conceived, turn on the question of whether sapience presupposes sentience (which is, as far as our understanding so far reaches, an exclusively biological phenomenon). John Searle's famous "Chinese Room" thought experiment pumps intuitions in this conceptual vicinity.

II. Representational and Propositional Dimensions of Practical and Discursive Intentionality

Within the general area marked out by the term 'intentionality', there are two distinctions it is important to keep in mind: the distinction between *practical* and *discursive* intentionality, and the distinction between *propositional* and *representational* intentionality. *Practical* intentionality is the sort of directedness at objects that animals exhibit when they deal skillfully with their world: the way a predator is directed at the prey it stalks, or the prey is directed at the predator it flees. It is a phenomenon of sentience, with the role objects, events, and situations play in the lived life of an animal providing the practical significances (food, threat…) that can be perceptually afforded. At the most abstract level of description, however, biological practical intentionality is an instance of a kind of broadly teleological directedness at objects that also has non-sentient examples. For any process that has a Test-Operate-Test-Exit feedback-loop structure, where operations on an object are controlled by information about the results of previous operations on it that are repeated until a standard is satisfied, can be seen as in a distinctive way "directed at" the objects the system both operates on and is informed about. This genus includes both finite-state automata executing conditional branched-schedule algorithms, for instance, in a radar-guided tracking anti-aircraft missile, and the fly-wheel governors that regulated the boiler-pressure of the earliest steam engines. *Discursive* intentionality is that exhibited by concept users in the richest sense: those that can make judgments or claims that are *about* objects in the semantic sense. The paradigm of the sort of sapience I am calling "discursive intentionality" is exhibited only by *language* users: ones who can *say* what they are thinking and talking about.

The distinction between *representational* and *propositional* intentionality is that between two dimensions of content intentional states can exhibit, corresponding to two of Searle's questions, quoted above: "What is S *of*? What is it an S *that*?". The answer to the first sort of question is the specification of an object represented by the state ("It is a belief of or about ships, shoes, sealing-wax…"), while the answer to the second sort of question is the specification of

what is believed or thought ("It is the belief that ships should be sea-worthy, that shoes are useful, that sealing-wax is archaic…"). The first expresses what we are thinking or talking *about*, and the second what we are thinking or saying (about it).

This distinction of two dimensions of contentfulness applies both to the practical and to the discursive species of intentionality. The dog believes *that* his master is home, and he believes that *of* Ben, his master. The principled difficulties we have with using the terms appropriate to discursive intentionality to specify precisely the propositional contents exhibited in practical intentionality (the dog does not really have the concepts specified by "master" and "home"—since it does not grasp most of the contrasts and implications essential to those concepts) do not belie the fact there is some content to his beliefs about that human, Ben, in virtue of which his belief that his master is about to feed him differs from his belief that his master is home, or that someone else will feed him.

Two opposed orders of explanation concerning the relations between practical and discursive intentionality are *pragmatism* and *platonism*. Pragmatism is the view that *discursive* intentionality is a species of *practical* intentionality: that knowing-*that* (things are thus-and-so) is a kind of knowing-*how* (to do something). What is *explicit* in the form of a principle is intelligible only against a background of *implicit* practices. The converse order of explanation, which dominated philosophy until the nineteenth century, is a kind of *intellectualism* that sees every implicit cognitive skill or propriety of practice as underwritten by a rule or principle: something that is or could be made discursively explicit. A contemporary version of *platonism* is endorsed by the program of symbolic artificial intelligence, which seeks to account for discursive intentionality as a matter of manipulating symbols according to definite rules. A contemporary version of *pragmatism* is endorsed by the program of pragmatic artificial intelligence, which seeks to account for discursive intentionality by finding a set of nondiscursive practices (practices each of which can be exhibited already by systems displaying only practical intentionality) that can be algorithmically elaborated into autonomous discursive practices.[8] Pragmatism need not take the reductive form of pragmatic AI, however.

What about the explanatory priority of the representational and propositional dimensions of intentionality? Here, too, various strategies are available. My own approach is to give

[8] I discuss these programs in more detail in Chapter Three of *Between Saying and Doing: Towards an Analytic Pragmatism* [Oxford University Press, 2008].

different answers depending on whether we are talking about practical or discursive intentionality. **Within *practical* intentionality, the *propositional* dimension should be understood in terms of the *representational* dimension. Within *discursive* intentionality, the *representational* dimension should be understood in terms of the *propositional*.** (Notice that the possibility of such a view would not even be visible to a theorist who did not make the distinctions with which I began this section.) The sort of representation that matters for understanding practical intentionality is the mapping relation that skillful dealings produce and promote between items in the environment and states of the organism. The usefulness of map representations depends on the goodness of inferences from map-facts (there is a blue wavy line between two dots here) to terrain-facts (there is a river between these two cities). The propositional content of the map-facts is built up out of representational relations that are sub-propositional (correlating blue lines and rivers, dots and cities). Such relations underwrite the representation-to-proposition order of explanation at the level of practical intentionality.

The considerations that speak for this order of explanation for *practical* intentionality are sometimes thought to speak for the same order of explanation for *discursive* intentionality. And the case could only get stronger when one conjoins that commitment with a pragmatist order of explanation relating practical and discursive intentionality. Nonetheless, I think there are strong reasons to endorse the explanatory priority of the propositional to the representational dimensions of intentionality at the level of discursive intentionality. They derive to begin with from consideration of the essentially *normative* character of discursive intentionality.

III. The Normativity of Discursive Intentionality

Kant initiated a revolution in thought about discursive intentionality. His most fundamental idea is that judgments and intentional doings are distinguished from the responses of nondiscursive creatures in that they are things the subject is in a distinctive way *responsible* for. They express *commitments*, or *endorsements*, they are exercises of the *authority* of the subject. Responsibility, commitment, endorsement, authority—these are all *normative* concepts. In undertaking a theoretical or practical discursive commitment that things are or shall be thus-andso, the knower/agent binds herself by rules (which Kant calls "concepts") that determine *what* she thereby becomes responsible for. For instance, in making the judgment that the coin is copper, the content of the concept copper that the subject applies determines that she is committed (whether she knows it or not) to the coin's conducting electricity, and melting at 1085° C., and that she is precluded from entitlement to the claim that it is less dense than water.

The difference between discursive and nondiscursive creatures is not, as Descartes had though, an *ontological* one (the presence or absence of some unique and spooky sort of mind-stuff), but a *deontological*, that is, normative one: the ability to bind oneself by concepts, which are understood as a kind of *rule*. Where the pre-Kantian tradition had focused on our grip on concepts (is it clear, distinct, adequate?), Kant focuses on their grip on us (what must one do to subject oneself to a concept in the form of a rule?). He understands discursive creatures as ones who live, and move, and have their being in a *normative* space.

The tradition Kant inherited pursued a bottom-up order of semantic (they said "logical") explanation that began with concepts, particular and general, representing objects and properties. At the next level, they considered how these representations could be combined to produce judgeable propositions of different forms ("Socrates is a man" "All men are mortal"). To the "doctrine of concepts" supporting the "doctrine of judgments" they then appended a "doctrine of syllogisms", which classified inferences as good or bad, depending on the kinds of judgments they involved. ("Socrates is a man, and all men are mortal, *so* Socrates is mortal.") This classical

theory was a paradigm of the order of explanation that proceeds from the *representational* to the *propositional* dimensions of discursive intentionality. In a radical break with tradition, Kant starts elsewhere. For him the fundamental intentional unity, the minimal unit of experience in the sense of sapient awareness is the *judgment* (proposition). For that is the minimal unit of *responsibility*. Concepts are to be understood top-down, by analyzing judgments (they are, he said "functions of judgment," rules for judging), looking at what contribution they make to the responsibilities undertaken by those who bind themselves by those concepts in judgment (and intentional agency). He initiated an order of explanation that moves ***from*** the *propositional* ***to*** the *representational* dimensions of discursive intentionality.

Pursuing that order of explanation in the context of his normative understanding of the *propositional* dimension of discursive intentionality led Kant to a normative account also of the *representational* dimension of discursive normativity. On the propositional side, the concept one has applied in judgment determines what one has made oneself responsible *for*. On the representational side, it determines what one has made oneself responsible *to*, in the sense of what sets the standard for assessments of the *correctness* of judgment. Kant sees that to treat something as a represent*ing*, as at least purporting to present something represent*ed*, is to acknowledge the *authority* of what is represented over assessments of the correctness of that representing. Discursive representation, too, is a normative phenomenon. And it is to be understood ultimately in terms of the contribution it makes to the normativity characteristic of propositional discursive intentionality.

Contemporary philosophical analyses of the normativity characteristic of discursive intentionality, along both propositional and representational dimensions, fall into two broad classes: *social-practical* and *teleosemantic*. Both are broadly *functionalist* approaches, in the sense that they look to the role discursive intentional states play in some larger system in explaining the norms they are subject to. Teleosemantic theories derive norms (what ought to follow, how the representing ought to be) from selectional, evolutionary, adaptive explanations of the advent of states and expressions that count as intentionally contentful (typically not just in the discursive, but also the practical sense) just in virtue of being governed by those norms. Ruth Millikan, for instance, defines Proper Function as that function that selectionally (counterfactually) explains the persistence of a feature or structure, in the sense that if such

features had not in the past performed that function, it would not have persisted.[9] Social practice theories date to Hegel, who accepted Kant's insight into the normative character of discursive intentionality, but sought to naturalize the norms in question (which Kant had transcendentalized). He understood normative statuses, such as commitment, entitlement, reponsibility, and authority, as instituted by practical normative attitudes. (Slogan: "All transcendental constitution is social institution."). On his account, genuine norms can only be instituted *socially:* as he put it, by "reciprocal recognition". The idea that discursive norms are to be understood as implicit in social practices was taken up from Hegel by the American pragmatists (C. S. Peirce, William James, and John Dewey), and later on by Ludwig Wittgenstein, who had independently discovered the normative character of discursive content.

The idea is that social norms are instituted when practitioners take or treat performances *as* appropriate or inappropriate, take or treat each other *as* committed, entitled, responsible, authoritative, and so on. The pragmatist thought is that even if the norms in question are discursive norms, adopting the instituting normative attitudes might require only practical intentionality. Practically punishing or rewarding performances is one way of treating them as inappropriate or appropriate. So for instance hominins in a certain tribe might practically treat it as inappropriate for anyone to enter a certain hut without displaying a leaf from a rare tree, by beating with sticks anyone who attempts to do so. In virtue of the role they play in this practice, the leaves acquire the practical normative significance of hut-licenses. In more sophisticated cases, the reward or punishment might itself be an alteration in normative status, regardless of its actual reinforcing effect. So one might treat a performance as appropriate by giving the performer a hut-license leaf, even if he has no interest in entering the hut.

[9] *Language, Thought, and other Biological Categories* [MIT Press, 1987].

IV. An Inferential Approach to Discursive Propositional Intentional Content

What makes something a specifically *discursive* norm? Discursive norms are norms governing the application of *concepts*, paradigmatically in judgment. Discursive norms govern the deployment of judgeable, that is, propositional intentional contents. In the context of a commitment to the *pragmatist* order of explanation, this question becomes: what kind of knowing *how* (to do something) amounts to knowing (or believing) *that* (things are thus-and-so)? What is the decisive difference—the difference that *makes* the difference—between a parrot who can reliably differentially respond to the visible presence of red things, perhaps by uttering "Rawk! That's red," on the one hand, and a human observer who can respond to the same range of stimuli by claiming and judging *that* something is red? What is it that the sapient, discursively intentional observer knows how to do that the merely sentient, practically intentional parrot does not?

The important difference is, to be sure, a matter of a distinctive kind of *understanding* that the concept-user evinces. The pragmatist wants to know: what practical abilities does that understanding consist in? We have acknowledged already the *normative* difference: the observer's performance does, as the parrot's does not, express an endorsement, the acknowledgement of a commitment. The key additional point to understand is that the content endorsed, the content the sapient observer is committed to qualifies as a *conceptual* content (of which specifically *propositional* contents are a principal species) just insofar as it is situated in a space of other such contents to which it stands in relations of material *consequence* and *incompatibility*. The observer knows how to make inferences and so draw conclusions from his commitment: to determine what *else* he has committed himself to by the claim that the apple is red (for instance, that it is colored, that it is ripe…). He knows how to distinguish what is *evidence* for and against that claim, and what else that commitment rules out as incompatible (for instance, that it is not wholly green). The sapient practically understands his commitment as taking up a stance in a network of related possible commitments, which stand to one another in

rational relations of material consequence and incompatibility. He is making a move in a practice of giving and asking for *reasons*, in which one move has normative consequences for what others are obligatory, permitted, or prohibited.

Material inferential (and incompatibility) relations, by contrast to *formal logical* inferential and incompatibility relations, articulate the contents of non-logical concepts. These are inferences such as "*A* is to the West of *B*, so *B* is to the East of *A*," "Lightning now, so thunder soon," and "If the sample is copper, then it will conduct electricity." Part of what one must do to count as understanding the contents of concepts such as East and West, lightning and thunder, copper and electrical conductor is to endorse inferences such as these. This is not to say that for each concept there is some meaning-constitutive set of material inferences one must endorse to count as understanding it. But if one makes *no* distinction, however partial and fallible, between material inferential and incompatibility relations that do and do not articulate the content of some concept, then one cannot count as a competent user of that concept.

Another way to get at the same point about the internal connection between *conceptual* contentfulness and *inferential* articulation is to consider the difference between *labeling* or *classifying* something and *describing* it. Any reliable differential responsive disposition imposes a classification on stimuli, distinguishing those that would from those that would not elicit a response of the given kind by the exercise of that reliable practical responsive capacity. The chunk of iron rusts in some environments and not others, the beam breaks under some loads and not others, the parrot squawks "Red!" in some situations and not others. What more is needed for such a performance to count not just as discriminating or labeling what elicits it, but also as describing it *as* red? The philosopher Wilfrid Sellars, one of my particular heroes, offers the following inferentialist answer:

> It is only because the expressions in terms of which we describe objects…locate these objects in a space of implications, that they describe at all, rather than merely label.[10]

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*

[10] 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.

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 inferential *consequences*, either immediately practical ("to be discarded/examined/kept") or for further classifications.

The inferentialist semantic claim is that what distinguishes specifically *discursive* (paradigmatically, but not exclusively, propositional) commitments is that their contents are articulated by the roles they play in material inferential and incompatibility relations. Grasping or understanding such contents is a kind of practical know-how: distinguishing in practice what follows from a given claimable or judgeable content, what it follows from, what would be evidence for it or against it, and what it would be evidence for or against. The practical inferential abilities to acknowledge the consequences of one's commitments for further commitments (both those one is committed to and those one is precluded from) and to distinguish evidence that would and would not entitle one to those commitments are what distinguish sapients from mere sentients, creatures that exhibit discursive intentionality from those that exhibit only practical intentionality.

V. The Relation of Language and Thought in Discursive Intentionality

It is obvious that there can be *practical* intentionality without language. Can there be *discursive* intentionality in the absence of language? Modern philosophers from Descartes through Kant took it also to be obvious that propositionally contentful thoughts and beliefs both antedate and are intelligible apart from their linguistic expression, which they understood in terms of symbols whose meanings are inherited from those antecedent prelinguistic discursive states and episodes. More recently, H. P. Grice extended this tradition, by understanding linguistic meaning in terms of speaker's meaning, and speaker's meaning in terms of the intention of a speaker to induce a belief in the audience by an utterance accompanied by the audience's recognition that the utterance was produced with that very intention. Another prominent line of thought in the area, due to Jerry Fodor, is the claim that public language is made possible by a language of thought, much of which is innate and so does not need to be learned.

A contrary order of explanation, identified with Wittgenstein among many others, gives explanatory priority to linguistic social practices in understanding discursive intentionality. Michael Dummett forcefully expresses one of the consequences of this approach:

> We have opposed throughout the view of assertion as the expression of an interior act of judgment; judgment, rather, is the interiorization of the external act of assertion.[11]

This way of turning the traditional explanatory strategy on its head is more extreme than is needed to acknowledge the crucial role of public language. Donald Davidson claims that to be a believer in the discursive sense one must be an interpreter of the speech of others.
But he also claims that:

> Neither language nor thinking can be fully explained in terms of the other, and neither has conceptual priority. The two are, indeed, linked in the sense that each

[11] *Frege's Philosophy of Language* [New York: Harper and Row, 1973], p. 362.

> requires the other in order to be understood, but the linkage is not so complete that either suffices, even when reasonably reinforced, to explicate the other.[12]

Although Davidson shares some important motivations with Dummett's purely linguistic theory, in fact these two views illustrate an important difference between two ways in which one might give prominence to linguistic practice in thinking about discursive intentionality. Davidson's claim, by contrast to Dummett's, serves to epitomize a *relational* view of the significance of language for sapience: taking it that concept use is not intelligible in a context that does not include language use, but not insisting that linguistic practices can be made sense of without appeal at the same time to intentional states such as belief.

According to such relational views, the transition from mere sentience to sapience (from practical to discursive intentionality) is effected by coming into language: coming to participate in discursive, social, linguistic practices. The capacity to *think* in the discursive sense—that is, to have propositionally or conceptually contentful thoughts, to be able to think that things are thus-and-so (a matter of knowing *that*, not just knowing *how*)—and the capacity to *talk* arise and develop together. For Wittgenstein, the essentiality of public language to the capacity for individual thought is a consequence of the normativity of discursive intentionality. He endorsed a pragmatist order of explanation that understands discursive norms as in the first instance implicit in social practices ("uses, customs, institutions" as he put it).[13] The capacity to make propositionally explicit claims and have conceptually contentful thoughts is intelligible only in the context of implicitly normative social linguistic practices.

---

[12] "Thought and Talk," in *Inquiries Into Truth and Interpretation* [New York: Oxford University Press, 1984], p. 156.

[13] Ludwig Wittgenstein *Philosophical Investigations*, G.E.M Anscombe [Wiley-Blackwell, 3rd edition, 1991] §199.

VI. Putting Together a Social Normative Pragmatics and an Inferential Semantics for Discursive Intentionality

An inferentialist about discursive content who understands discursive norms as implicit in social linguistic practice and holds a relational view of the priority of language and thought will take it that the core of discursive intentionality is to be found in the role *declarative sentences* play in expressing *propositional* contents in speech acts of *assertion*. This connection between the *syntactic* category of declarative sentences, the *semantic* category of propositions, and the *pragmatic* category of assertions is the *iron triangle of discursiveness*. A pragmatist about the relations between them takes it that the syntactic and semantic elements are ultimately to be understood in terms of the pragmatic one. It is their role in the practice of assertion, of claiming that things are thus-and-so, that is appealed to in picking out declarative sentences and propositional contents. Propositional contents are what can both serve as and stand in need of *reasons*—that is, can perform the office both of premise and of conclusion in *inferences*. So the inferentialist pragmatist takes it that what distinguishes the speech act of assertion is its role in practices of giving and asking for *reasons*.

One way of putting together a social normative pragmatics and an inferential semantics for discursive intentionality is to think of linguistic practices in terms of *deontic scorekeeping*. Normative statuses show up as social statuses. The paradigmatic deontic status is *commitment*. The idea is that we should understand what one is doing in making an assertion is undertaking a distinctive kind of *commitment*: *making* a claim is *staking* a claim. If acquiring the status of being committed in the way standardly undertaken by assertively uttering the sentence *p* is to be significant, it must have consequences. The inferentialist says to look for *inferential* consequences (and antecedents): what *else* one becomes committed to by asserting *p* (what follows from *p*) and what would commit one to it (what it follows from). The pragmatist says to understand that in terms of what one is obliged (or permitted) to *do*, upon asserting *p*. To understand an assertional speech act is to know how to *keep score* on the commitments the speaker has undertaken by peforming that act. In undertaking commitment to *p*, the asserter has *obliged* herself to acknowledge *other* commitments: those that follow from it. She has also

authorized other interlocutors to attribute that commitment to her. Further, she has obliged herself to offer a *justification* (give reasons) for the claim, if her authority is suitably challenged. The idea is that exercising such inferentially articulated *authority* and fulfilling such inferentially articulated *responsibility* is what one must *do* (the task responsibilities one must carry out) in order to count as responsible for or committed—not now to do something, but to what in this social-practical scorekeeping context shows up as the propositional content *p*.

For such an idealized assertional practice to count as one of giving and asking for reasons, there must be a difference between commitments for which one *can* give a reason (so fulfilling one's justificatory task-responsibility) and those for which one can*not*. That is, there must be a distinction between commitments to which an asserter is (rationally, inferentially, by one's evidence) *entitled*, and those to which the assertor is not entitled. So in practice to take or treat a performance *as* an assertion of a particular propositional content, other interlocutors must keep track not only of how that performance changes the score of what the asserter is committed to, but also what she (and others) are entitled to. Discursive scorekeeping requires attributing two sorts of deontic status: commitments and entitlements (to commitments), and knowing how different speech acts change the deontic "scores" of various interlocutors—who may become entitled to new commitments by relying on the authority of other asserters (to whom they can then defer their justificatory responsibility). This deontic scorekeeping story is a sketch of how discursive intentionality is intelligible as emerging from exercises of practical intentionality that have the right normative and social structure.[14]

Scorekeepers acknowledging and attributing two kinds of normative deontic status, commitments and entitlements, can distinguish three kinds of practical consequential relations among them, which generate three flavors of inferential relations, and a relation of material incompatiblity. Scorekeepers who take anyone who is entitled to *p* to be (prima facie) entitled to *q* thereby practically endorse a *permissive* inferential (probatively evidential) relation between *p* and *q*. This is a generalization, from the formal-logical to the contentful material case, of *inductive* inference. (The barometer is falling, so there will be a storm.) Scorekeepers who take anyone who is committed to *p* to be committed to *q* thereby practically endorse a *committive*

[14] I develop this model further in *Making It Explicit: Reasoning, Representing, and Discursive Commitment* [Harvard University Press, 1994]—especially Chapter Three.

inferential (dispositive evidential) relation between *p* and *q*. This is a generalization, from the formal-logical to the contentful material case, of *deductive* inference. (If the sample is pure copper, it will conduct electricity.) Scorekeepers who practically take or treat anyone who is committed to *p not* to be entitled to *q*, and vice versa, thereby treat the two claims they express as materially *incompatible*. (The plane figure cannot be both square and circular.) Scorekeepers for whom everything incompatible with *q* is incompatible with *p* thereby practically take or treat *q* as *incompatibility-entailed* by *p*. (Everything incompatible with Pedro being a mammal is incompatible with Pedro being a donkey, so his being a donkey in this sense entails his being a mammal.) These are modally robust, counterfactual-supporting entailments.

When an interlocutor makes an assertion by uttering *p*, scorekeepers take or treat him as also committed to committive consequences of *p*, withdraw attributed entitlements to any claims incompatible with *p*, and if they take it that he is also entitled to *p*, attribute further entitlements to its permissive consequences to him and to anyone in the audience not precluded by virtue of incompatible commitments. Adopting these practical deontic scorekeeping attitudes is what those who appreciate the practical significance of the speech act must *do* in order thereby to count as implicitly taking or treating the utterance as playing the functional role in virtue of which it expresses a *propositional* discursive content. Other uses of language are built on this assertional-inferential core (the "downtown" of language), and make use of the conceptual contents conferred by it.

VI. Logic: the Organ of Semantic Self-Consciousness

This inferentialist, social practical story about the structure of *practical* intentionality (knowing *how*, abilities) that adds up to *discursive* intentionality (knowing or believing *that* things are thus-and-so) presents an approach to *pragmatics*, or the *use* of language: the norms implicit in scorekeeping practices. According to that story, expressions come to have *propositional discursive semantic* content, and so are able to make something explicit,in the sense of its being sayable, claimable, thinkable, in virtue of being practically taken or treated *as* standing in relations of material inference-and-incompatibility. Building on this kind of basic discursive (sapient) intentional practices and abilities, it is also possible for such practitioners to make propositionally explicit those normative material inferential and incompatibility relations, which are initially implicit in the practical attitudes discursive scorekeepers adopt to one another.

Most centrally, inferential (including *material* inferential) relations can be put in claimable (propositional, explicit) form by the use of *conditional* locutions. One can explicitly express one's endorsement of the inference from *p* to *q* by asserting "If *p* then *q*." Incompatibility relations can be made explicit using *negation* operators. One can explicitly express one's taking *p* to be incompatible with *q* by asserting "Not (*p*&*q*)."[15] Conditional and negation operators are *logical* vocabulary. (Indeed, versions of them suffice to define the classical propositional calculus.) The expressive role characteristic of logical vocabulary is to make explicit the material inferential and incompatibility relations in virtue of which non-logical vocabulary expresses the semantic content that it does. It is by playing the role they do in a network of such relations that expressions acquire the propositional content that makes possible the discursive, sapient awareness that consists in explicitly claiming or judging *that* things are thus-and-so. Logical vocabulary makes possible explicit, discursive, sapient awareness of those

[15] I expound this approach to logic in more detail in "From Logical Expressivism to Expressivist Logic: Sketch of a Program and Some Implementations" Nous: Philosophical Issues, Volume 28, Issue 1, October 2018, (a volume devoted to the philosophy of logic), pp. 70-88.

very semantogenic material inferential and incompatibility relations. Logic is the organ of semantic self-consciousness.

On this account of the expressive role that demarcates vocabulary as distinctively logical, it is intelligible that there should be creatures that are *rational*, but not yet *logical*. To be rational is to engage in practices of giving and asking for reasons, that is, making inferentially articulated assertions and justifying them. To do that one must attribute and acknowledge commitments and entitlements, and practically keep track of their inferential relations along all three dimensions those two deontic statuses generate: permissive, committive, and incompatibility entailments. But one need not yet deploy specifically logical vocabulary, which permits one to make explicit and so be discursively aware of those material inferential and incompatibility relations. In being rational, one already knows how to do everything one needs to know how to do to introduce logical vocabulary. But until such semantically explicitating vocabulary actually is deployed, rational creatures need not be semantically self-conscious, that is, logical creatures. We are not like that. But our hominin ancestors might have been.

VII. Pragmatic Social Normative Perspectives and the Representational Dimension of Discursive Semantic Content

Practically keeping track of inferentially-articulated commitments and entitlements (that is, engaging in discursive practices) requires distinguishing between the normative statuses one *attributes* (to another) and those one *acknowledges* (oneself). This distinction of social perspective between normative attitudes means that there are two points of view from which one can assess another's consequential commitments. For the auxiliary hypotheses or collateral premises one conjoins to another's avowed commitment to extract its consequences (whether permissive, committive, or incompatibility-entailed) can be drawn either from other commitments one *attributes* to that interlocutor, or from those one *undertakes* oneself. Suppose S attributes to A commitment to the claim "Benjamin Franklin was a printer," (perhaps on the basis of hearing A make that assertion). If S also attributes to A commitment to "Benjamin Franklin is (=) the inventor of the lightning rod." Then S should also attribute to A commitment to "The inventor of the lightning rod was a printer." But suppose S, but not A, is committed to "Benjamin Franklin is (=) the inventor of bifocals." Should S attribute to A commitment to "The inventor of bifocals was a printer"? Given the fact (as S takes it) that Franklin invented bifocals, that is indeed a consequence of A's original claim. In the context of that fact, a claim about Ben Franklin is a claim about the inventor of bifocals, whether or not A realizes that. So in a genuine and important sense, A has, without knowing it, committed herself to the inventor of bifocals having been a printer. But that is a *different* sense from that in which A has committed herself to the inventor of the lightning rod having been a printer.

When the practical adoption of a normative attitude of attributing a commitment to another interlocutor is made propositionally explicit by the use of locutions that let one *say* what commitments one practically attributes to another, this difference in social perspective manifests itself in two different kinds of ascription of propositional attitude. Consequential commitments attributed solely on the basis of commitments the target would assert are ascribed *de dicto*. S can say "A claims (believes, is committed to the claim) *that* the inventor of the lightning rod was a

printer." Consequential commitments attributed partly on the basis of commitments the target would assert and partly by the use of collateral premises that the *attributor*, but *not* the target of the attribution, would assert are ascribed *de re*. S can say "A claims *of* the inventor of bifocals that he was a printer. In putting things this way, S marks that while he is *attributing* to A responsibility for the overall claim, S is himself *undertaking* responsibility for the substitution inference licensed by the identity "Benjamin Franklin is the inventor of bifocals," (commitment to which he does *not* attribute to A).[16]

Propositional attitude ascribing locutions, such as "claims" and "believes" let their users make explicit their practical normative scorekeeping attitudes of attributing commitments, that is, using such vocabulary empowers them to *say that* they adopt such attitudes, which otherwise remain implicit in what they practically do. Performing this expressive office with respect to social normative attitudes, on the side of pragmatics, marks them as another species of the same explicitating genus as logical vocabulary, which does corresponding service on the semantic side, by making explicit inferential commitments. What S is *doing* in making *de re* ascriptions is expressing the distinction of social perspective between commitments *attributed* (Ben Franklin was a printer) and those *undertaken* (Ben Franklin invented bifocals). But what one is *saying* is what the one to whom the commitments are ascribed was talking *about*. *De re* ascriptions of propositional attitude are the home language-game of *representational* locutions: the ones used to make explicit what one is talking or thinking *of* or *about*. What they make explicit is the representational dimension of discursive intentionality.

That representational dimension is always already implicit in the distinction of social perspective that is integral to keeping track of others' inferentially articulated commitments. For discursive deontic scorekeepers, players of the game of giving and asking for reasons, care about what follows from others' claims for two reasons. They care about the consequential commitments that would be ascribed *de dicto* because they want to know what else the target would endorse, and what she will do based on the commitments she acknowledges. They care about the consequential commitments that would be ascribed *de re* because they want to extract *information* from the claims of others—that is, claims that the attributor can use as premises in

[16] I discuss the distinction between propositional attitude ascriptions *de dicto* and *de re* in Chapter Eight of *Making It Explicit* [op. cit.], and Chapter Three of *Tales of the Mighty Dead: Historical Essays in the Metaphysics of Intentionality* [Harvard University Press, 2002].

his own inferences. If S attributes to A the intention to shoot a deer, and the belief that the tawny creature in front of her *is* a deer, the *de dicto* ascription "A believes that the tawny creature in front of her is a deer, the shooting of which would fulfill her intention," S will predict that A will shoot. If S, but not A, believes that the tawny creature in front of A is (=) a cow, then S's *de re* ascription "A believes *of* the cow in front of her *that* it is a deer, the shooting of which would fulfill her intention," S will predict that the result of A's action will be the shooting of a cow. That is an inference that S is in a position to extract from A's avowed commitments, even though that information is not available to A. Keeping track of what premises are available for the reasoning of others and what premises are available for our own reasoning is what we are doing when we talk or think about what we are talking or thinking *about*. These *de re* ascriptions of propositional attitude make explicit the *representational* dimension of discursive intentionality. I have suggested how it can be understood in terms of the social normative inferential articulation of the more basic *propositional* dimension of discursive intentionality.

END of Lecture 2

Semantic Inferentialism and Logical Expressivism[17]

I: Introduction

In this essay I want to introduce a way of thinking about semantics that is different from more familiar ones, and on that basis also a new way of thinking about logic. In case that seems insufficiently ambitious, I'll introduce these ideas by sketching a different way of thinking about some important episodes in the history of philosophy, in the era that stretches from Descartes to Kant. I'm going to explain and motivate the two ideas indicated in the title by putting together considerations drawn from three different thinkers: Frege, Dummett, and Sellars or, as I think of them: the sage of Jena, the sage of Oxford, and the sage of Pittsburgh. In each case I'll be picking up strands other than those usually emphasized in reading these figures.

[17] This lecture is drawn from Chapter One of *Articulating Reasons: An Introduction to Inferentialism* [Harvard University Press, 2000].

II: Representationalism and Inferentialism

Pre-Kantian empiricists and rationalists alike were notoriously disposed to run together causal and conceptual issues, largely through insufficient appreciation of the normative character of the "order and connection of ideas" that matters for concepts. But there is another, perhaps less appreciated, contrast in play during this period, besides that of the causal and the conceptual, the origin and the justification of our ideas. Enlightenment epistemology was always the home for two somewhat uneasily coexisting conceptions of the conceptual. The fundamental concept of the dominant and characteristic understanding of cognitive contentfulness in the period initiated by Descartes is of course *representation*. However there is a minority semantic tradition that takes *inference* rather than representation as its master concept.

Rationalists such as Spinoza and Leibniz accepted the central role of the concept of representation in explaining human cognitive activity. But they were not prepared to accept Descartes' strategy of treating the possession of representational content as an unexplained explainer—just dividing the world into what is by nature a representing and what by nature can only be represented. Each of them developed instead an account of what it is for one thing to represent another, in terms of the *inferential* significance of the representing. They were explicitly concerned, as Descartes was not, to be able to explain what it is for something to be understood, taken, treated, or employed *as* a representing *by* the subject: what it is for it to be a representing *to* or *for* that subject (to be "tanquam rem", as if of things, as Descartes puts it). Their idea was that the way in which representings point beyond themselves to something represented is to be understood in terms of *inferential* relations among representings. States and acts acquire content by being caught up in inferences, as premises and conclusions.

Thus a big divide within Enlightenment epistemology concerns the relative explanatory priority accorded to the concepts of representation and inference. The British empiricists were more puzzled than Descartes about representational purport: the property of so much as seeming to be *about* something. But they were clear in seeking to derive inferential relations from the contents of representings, rather than the other way around. In this regard they belong to the still-dominant tradition that reads inferential correctnesses off from representational

correctnesses, which are assumed to be antecedently intelligible. That is why Hume could take for granted the contents of his individual representings, but worry about how they could possibly underwrite the correctness of inductive inferences. The post-Cartesian rationalists, the claim is, give rise to a tradition based on a complementary semantically reductive order of explanation. (So Kant, picking up the thread from this tradition, will come to see their involvement in counterfactually robust inferences as essential to empirical representations having the contents that they do.) These *inferentialists* seek to define representational properties in terms of inferential ones, which must accordingly be capable of being understood antecedently. They start with a notion of content as determining what is a *reason* for what, and understand truth and representation as features of ideas that are not only manifested in, but actually *consist* in their role in reasoning. I actually think that the division of pre-Kantian philosophers into representationalists and inferentialists cuts according to deeper principles of their thought than does the nearly coextensional division of them into empiricists and rationalists, though it goes far beyond my brief to argue for that thesis here.

III: Inferentialism and Noninferential Reports

The concepts for which inferential notions of content are least obviously appropriate are those associated with observable properties, such as colors. For the characteristic use of such concepts is precisely in making *non*inferential reports, such as "This ball is red." One of the most important lessons we can learn from Sellars' masterwork, "Empiricism and the Philosophy of Mind" (as from the *Sense Certainty* section of Hegel's *Phenomenology*) is the inferentialist one that even such noninferential reports must be inferentially articulated. Without that requirement, we can't tell the difference between noninferential reporters and automatic machinery such as thermostats and photocells, which also have reliable dispositions to respond differentially to stimuli. What is the important difference between a thermostat that turns the furnace on when the temperature drops to 60 degrees, or a parrot trained to say "That's red," in the presence of red things, on the one hand, and a genuine noninferential reporter of those circumstances, on the other? Each classifies particular stimuli as being of a general kind, the kind, namely, that elicits a repeatable response of a certain sort. In the same sense, of course, a chunk of iron classifies its environment as being of one of two kinds, depending on whether it responds by rusting or not. It is easy, but uninformative, to say that what distinguishes reporters from reliable responders is awareness. In this use, the term is tied to the notion of understanding--the thermostat and the parrot don't understand their responses, those responses mean nothing to them, though they can mean something to us. We can add that the distinction wanted is that between merely responsive classification and specifically *conceptual* classification. The reporter must, as the parrot and thermostat do not, have the *concept* of temperature or cold. It is classifying under such a concept, something the reporter understands or grasps the meaning of, that makes the relevant difference.

It is at this point that Sellars introduces his central thought: that for a response to have *conceptual* content is just for it to play a role in the *inferential* game of making claims and giving and asking for reasons. To grasp or understand such a concept is to have practical mastery over the inferences it is involved in--to know, in the practical sense of being able to distinguish (a kind of know-*how*), what follows from the applicability of a concept, and what it follows from. The parrot doesn't treat "That's red" as incompatible with "That's green", nor as following from

"That's scarlet" and entailing "That's colored." Insofar as the repeatable response is not, for the parrot, caught up in practical proprieties of inference and justification, and so of the making of further judgements, it is not a *conceptual* or a *cognitive* matter at all.

It follows immediately from such an inferential demarcation of the conceptual that in order to master *any* concepts, one must master *many* concepts. For grasp of one concept consists in mastery of at least some of its inferential relations to other concepts. Cognitively, grasp of just one concept is the sound of one hand clapping. Another consequence is that to be able to apply one concept *non*inferentially, one must be able to use others *inferentially*. For unless applying it can serve at least as a premise from which to draw inferential consequencences, it is not functioning as a concept at all. So the idea that there could be an autonomous language game, one that could be played though one played no other, consisting entirely of noninferential reports (in the case Sellars is most concerned with in *EPM*, even of the current contents of one's own mind) is a radical mistake. (Of course this is compatible with there being languages without theoretical concepts, that is, concepts whose *only* use is inferential. The requirement is that for *any* concepts to have reporting uses, some concepts must have *non*reporting uses.)

IV: Frege on Begriffliche Inhalt

My purpose at the moment, however, is not to pursue the *consequences* of the inferential understanding of conceptual contents that Sellars recommends, but its *antecedents*. The predecessor it is most interesting to consider is the young Frege. Frege may seem an unlikely heir to this inferentialist tradition. After all, he is usually thought of as the father of the contemporary way of working out the *representationalist* order of explanation, which starts with an independent notion of relations of reference or denotation obtaining between mental or linguistic items and objects and sets of objects in the largely nonmental, nonlinguistic environment, and determines from these in the familiar fashion, first truth conditions for the sentential representings built out of the subsentential ones, and then, from these, a notion of goodness of inference understood in terms of set-theoretic inclusions among the associated sets of truth conditions. But insofar as it is appropriate to read this twentieth century story back into Frege at all, and I am not sure that it is, it would be possible only beginning with the Frege of the 1890's. He starts his semantic investigations, not with the idea of reference, but with that of inference. His seminal first work, the *Begriffsschrift* of 1879, takes as its aim the explication of "conceptual content" [begriffliche Inhalt]. The qualification "conceptual" is explicitly construed in inferential terms:

2] ...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. The two propositions 'the Greeks defeated the Persians at Plataea' and 'the Persians were defeated by the Greeks at Plataea' differ in the former way; even if a slight difference of sense is discernible, the agreement in sense is preponderant. Now I call that part of the content that is the same in both the conceptual content. Only this has significance for our symbolic language [Begriffsschrift]... In my formalized language [BGS]...only that part of judgments which affects the possible inferences is taken into consideration. Whatever is needed for a correct ['richtig', usually misleadingly translated as 'valid'] inference is fully expressed; what is not needed is...not.[18]

Two claims have the same conceptual content iff they have the same inferential role: a good inference is never turned into a bad one by substituting one for the other. This way of specifying the explanatory target to which semantic theories, including referential ones, are directed is picked up by Frege's student Carnap,

[18] Frege, *Begriffsschrift* (hereafter *BGS*), section 3.

who in the Logical Syntax of Language defines the content of a sentence as the class of non-valid sentences which are its consequences (i.e. can be inferred from it). Sellars in turn picks up the idea from him, as his references to this definition indicate.

By contrast, the tradition Frege initiated in the 1890's makes truth, rather than inference, primary in the order of explanation. Dummett says of this shift:

3] ...in this respect (and [Dummett implausibly but endearingly hastens to add] in this respect alone) Frege's new approach to logic was retrograde. He characterized logic by saying that, while all sciences have truth as their goal, in logic truth is not merely the goal, but the object of study. The traditional answer to the question what is the subject-matter of logic is, however, that it is, not truth, but inference, or, more properly, the relation of logical consequence. This was the received opinion all through the doldrums of logic, until the subject was revitalized by Frege; and it is, surely, the correct view.[19]

~~And again:~~

~~4] It remains that the representation of logic as concerned with a characteristic of sentences, truth, rather than of transitions from sentences to sentences, had highly deleterious effects both in logic and in philosophy. In philosophy it led to a concentration on logical truth and its generalization, analytic truth, as the problematic notions, rather than on the notion of a statement's being a deductive consequence of other statements, and hence to solutions involving a distinction between two supposedly utterly different kinds of truth, analytic truth and contingent truth, which would have appeared preposterous and irrelevant if the central problem had from the start been taken to be that of the character of the relation of deductive consequence~~.[20]

---

[19] Dummett, *Frege's Philosophy of Language* [Harper & Row 1973] (hereafter *FPL*), p. 432.

[20] Dummett, *FPL*, p. 433. A few comments on this passage: First, the "deleterious effects in logic" Dummett has in mind include taking logics to be individuated by their theorems rather than their consequence relations. Although one can do things either way for classical logic, in more interesting cases logics can have the same theorems but different consequence relations. Second, the contrast with analytic is not obviously contingent—why rule out the possibility of necessity that is not conceptual, but, say, physical? Third, the closing claim seems historically wrong. Kant already distinguished analytic from synthetic judgments, and his concerns did not evidently stem from concern with the subject-matter of logic. I include the passage anyway, since I think the shift in emphasis Dummett is endorsing is a good one, although the reasons he advances need filling in and cleaning up.

The important thing to realize is that the young Frege has not yet made this false step.

Two further points to keep in mind regarding this passage are: first, shifting from concern with inference to concern with truth is one move, understanding truth in terms of prior primitive reference relations is another. Since the mature Frege treats truth as indefinable and primitive, the extraction of a representationalist commitment even from the texts of the 1890's requires further showing (compare Davidson's truth-without-reference view in our own day). Second, understanding the topic of logic in terms of inference is not the same as seeing it in terms of logical inference, or of "deductive consequence", as Dummett puts it (I'll talk about this below under the heading of "formalism" about inference). The view propounded and attributed to Frege below is different, and from the contemporary vantage-point, more surprising, than that Dummett endorses here.

V: Material Inference

The kind of inference whose correctnesses determine the conceptual contents of its premises and conclusions may be called, following Sellars, *material* inferences. As examples, consider the inference from "Pittsburgh is to the West of Princeton" to "Princeton is to the East of Pittsburgh", and that from "Lightning is seen now" to "Thunder will be heard soon". It is the contents of the concepts West and East that make the first a good inference, and the contents of the concepts lightning and thunder, as well as the temporal concepts, that make the second appropriate. Endorsing these inferences is part of grasping or mastering those concepts, quite apart from any specifically *logical* competence.

Often, however, *inferential* articulation is identified with *logical* articulation. Material inferences are accordingly treated as a derivative category. The idea is that being rational—being subject to the normative force of the better reason, which so puzzled and fascinated the Greeks—can be understood as a purely logical capacity. In part this tendency was encouraged by merely verbally sloppy formulations of the crucial difference between the inferential force of reasons and the physically efficacious force of causes, which render it as the difference between 'logical' and 'natural' compulsion. Mistakes ensue, however, if the concept <u>logical</u> is employed with these circumstances of application conjoined with consequences of application that restrict the notion of logical force of reasons to formally valid inferences. The substantial commitment that is fundamental to this sort of approach is what Sellars calls

5] ...the received dogma...that the inference which finds its expression in "It is raining, therefore the streets will be wet" is an enthymeme.[21]

According to this line of thought, wherever an inference is endorsed, it is because of belief in a conditional. Thus the instanced inference is understood as implicitly involving the conditional "If it is raining, then the streets will be wet". With that "suppressed" premise supplied, the inference is an instance of the formally valid scheme of conditional detachment.

[21] Sellars "Inference and Meaning," reprinted in *Pure Pragmatics and Possible Worlds* J. Sicha (ed.) [Ridgeview Publishing Co. 1980] (hereafter, *PPPW*), pp. 261/313.

The "dogma" expresses a commitment to an order of explanation that treats all inferences as good or bad solely in virtue of their form, with the contents of the claims they involves mattering only for the truth of the (implicit) premises. According to this way of setting things out, there is no such thing as material inference. This view, which understands "good inference" to mean "formally valid inference", postulating implicit premises as needed, might be called a formalist approach to inference. It trades primitive goodnesses of inference for the truth of conditionals. Doing so is taking the retrograde step that Dummett complains about. (It is also what introduces the problem Lewis Carroll exposes in "Achilles and the Tortoise.") The grasp of logic that is attributed must be an implicit grasp, since it need be manifested only in distinguishing material inferences as good and bad, not in any further capacity to manipulate logical vocabulary or endorse tautologies involving them. But what then is the explanatory payoff from attributing such an implicit logical ability rather than just the capacity to assess proprieties of material inference?

The approach Sellars endorses is best understood by reference to the full list of alternatives he considers:

6] ...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".[22]

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:

7] ...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.[23]

Should inferentialist explanations begin with inferences pertaining to propositional *form*, or those pertaining to propositional *content*? One important consideration is that the notion of formally valid inferences is definable in a natural way from that of materially correct ones, while there is no converse route. For given a subset of vocabulary that is privileged or distinguished somehow, an inference can be treated as good in virtue of its form, with respect to that vocabulary, just in case

--it is a materially good inference and

-- it cannot be turned into a materially bad one by substituting non-privileged for non-privileged vocabulary, in its premises and conclusions.

Notice that this substitutional notion of formally good inferences need have nothing special to do with *logic*. If it is *logical* form that is of interest, then one must antecedently be able to distinguish some vocabulary as peculiarly logical. That done, the Fregean semantic strategy of looking for inferential features that are invariant under substitution yields a notion of *logically* valid inferences. But if one picks out *theol*ogical (or aesthetic) vocabulary as privileged, then looking at which substitutions of non-theological (or non-aesthetic) vocabulary for non-theological (non-aesthetic) vocabulary preserve material goodness of inference will pick out

[22] Sellars, "Inference and Meaning" *PPPW* pp. 265/317.
[23] Sellars, "Inference and Meaning" *PPPW* pp. 284/336.

inferences good in virtue of their *theol*ogical (or aesthetic) form. According to this way of thinking, the *formal* goodness of inferences derives from and is explained in terms of the *material* goodness of inferences, and so ought not to be appealed to in explaining it. Frege's inferentialist way of specifying the characteristic linguistic role in virtue of which vocabulary qualifies as logical is discussed below.

VI: Elucidative Rationality

So far I have indicated briefly two related claims: that conceptual contents are inferential roles, and that the inferences that matter for such contents in general must be conceived to include those that are in some sense *materially correct*, not just those that are *formally valid*. I'll argue in a moment that a commitment to the second of these, no less than the first, is to be found already in Frege's early writings, though not in the developed form to which Sellars brings it. But in both thinkers these ideas are combined with a third, which I believe makes this line of thought especially attractive. In one of his early papers, Sellars introduces the idea this way:

8] Socratic method serves the purpose of making explicit the rules we have adopted for thought and action, and 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'.[24]

Sellars understands such modal statements as inference licenses, which formulate as the content of a claim the appropriateness of inferential transitions. More than this, he understands the function of such statements to be making explicit, in the form of assertible rules, commitments that had hitherto remained implicit in inferential practices. Socratic method is a way of bringing our practices under rational control, by expressing them explicitly in a form in which they can be confronted with objections and alternatives, a form in which they can be exhibited as the conclusions of inferences seeking to justify them on the basis of premises advanced as reasons, and as premises in further inferences exploring the consequences of accepting them.

In the passage just quoted, Sellars tells us that the enterprise within which we ought to understand the characteristic function of inference licenses is a form of rationality that centers on the notion of *expression*: making *explicit* in a form that can be thought or said, what is *implicit* in what is done. This is a dark and pregnant claim, but I believe it epitomizes a radical and distinctive insight. In what follows I hope to shed some light on it and its role in an inferentialist vision of things. The general idea is that the paradigmatically rational process that Sellars

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

invokes under the heading of "Socratic method" depends upon the possibility of making implicit commitments explicit in the form of claims. *Expressing* them in this sense is bringing them into the game of giving and asking for reasons as playing the special sort of role in virtue of which something has a conceptual content at all: namely an inferential role, as premise and conclusion of inferences. This sort of rationality is distinct from, but obviously related to the sort of rationality that then consists in making the appropriate inferential moves. Even totalitarian versions of the latter, for instance those that would assimilate all goodness of inference to logical validity, or to instrumental prudence (that is, efficiency at getting what one wants), depend upon the possibility of expressing considerations in a form in which they can be given as reasons, and reasons demanded for them. All the more does Socratic reflection on our practices, particularly on those material-inferential practices that determine the conceptual contents of thoughts and beliefs, depend on the possibility of their explicit expression.

VII: Frege on the Expressive Role of Logic

To begin to explicate this notion of explication, it is helpful to return to the consideration of the young Frege's inferentialist program. Frege's *Begriffsschrift* is remarkable not only for the inferential idiom in which it specifies its topic, but equally for how it conceives its relation to that topic. The task of the work is officially an expressive one; not to prove something, but to say something. Frege's logical notation is designed for expressing conceptual contents, making explicit the inferential involvements that are implicit in anything that possesses such content. As passage [2] quoted above puts it: "Whatever is needed for a correct inference is fully expressed". Talking about this project, Frege says:

9] Right from the start I had in mind the expression of a content...But the content is to be rendered more exactly than is done by verbal language... Speech often only indicates by inessential marks or by imagery what a concept-script should spell out in full.[25]

The concept-script is a formal language for the explicit codification of conceptual contents. In the Preface to BGS, Frege laments that even in science concepts are formed haphazardly, so that the ones employing them are scarcely aware of what they mean, of what their content really is. When the correctness of particular inferences is at issue, this sort of unclarity may preclude rational settlement of the issue. What is needed is a notation within which the rough-and-ready conceptual contents of the sciences, beginning with mathematics, can be reformulated so as to wear their contents on their sleeves. The explanatory target here avowedly concerns a sort of inference, not a sort of truth, and the sort of inference involved is content-conferring material inferences, not the derivative formal ones.

Frege explicitly contrasts his approach with that of those, such as Boole, who conceive their formal language only in terms of formal inference, and so express no material contents:

10] The reason for this inability to form concepts in a scientific manner lies in the lack of one of the two components of which every highly developed language must consist. That is, we may distinguish the formal part...from the material part proper. The

[25] Frege, from "Boole's logical Calculus and the Concept-script", *Posthumous Writings* (hereafter *PW*) pp.12-13.

signs of arithmetic correspond to the latter. What we still lack is the logical cement that will bind these building stones firmly together...In contrast, Boole's symbolic logic only represents the formal part of the language.[26]

By contrast:

11] 1. My concept-script has a more far-reaching aim than Boolean logic, in that it strives to make it possible to present a content when combined with arithmetical and geometrical signs...
2. Disregarding content, within the domain of pure logic it also, thanks to the notation for generality, commands a somewhat wider domain...
4. It is in a position to represent the formation of the concepts actually needed in science...[27]

It is the wider domain to which his expressive ambition extends that Frege sees as characteristic of his approach. Since contents are determined by inferences, expressing inferences explicitly will permit the expression of any sort of content at all:

12] It seems to me to be easier still to extend the domain of this formula language to include geometry. We would only have to add a few signs for the intuitive relations that occur there...The transition to the pure theory of motion and then to mechanics and physics could follow at this point.[28]

Frege's early understanding of logic offers some specific content to the notion of explicitly expressing what is implicit in a conceptual content, which is what is required to fill in a notion of expressive or elucidating rationality that might be laid along side (and perhaps even be discovered to be presupposed by) notions of rationality as accurate representation, as logically valid inference, and as instrumental practical reasoning. Before the fateful step from seeing logic as an attempt to codify inferences to seeing it as the search for a special kind of truth is made, which Dummett bemoans, Frege's aim is to introduce vocabulary that will let one *say* (explicitly) what otherwise one can only *do* (implicitly). Consider the conditional, with which the *Begriffsschrift* begins. Frege says of it:

13] The precisely defined hypothetical relation between contents of possible judgments [Frege's conditional] has a similar significance for the foundation of my concept-script to that which identity of extensions has for Boolean logic.[29]

[26] Frege, *PW* p. 13.
[27] Frege, *PW* p. 46.
[28] Frege, *Begriffsschrift* Preface, in van Heijenoort (ed.) *From Frege to Godel* Harvard Press, 1967 p. 7.
[29] Frege, *PW* p. 16.

[I think it is hard to overestimate the importance of this passage in understanding what is distinctive about Frege's *Begriffsschrift* project. After all, contemporary Tarskian model-theoretic semantics depends precisely on relations among extensions. Frege is saying that his distinctive idea—in what is, after all, the founding document of modern formal logic—is to do things otherwise.] Why the conditional? Prior to the introduction of such a conditional locution, one could *do* something, one could treat a judgement as having a certain content (implicitly attribute that content to it) by endorsing various inferences involving it and rejecting others. After conditional locutions have been introduced, one can *say*, as part of the content of a claim (something that can serve as a premise and conclusion in inference), *that* a certain inference is acceptable. One is able to make explicit material inferential relations between an antecedent or premise and a consequent or conclusion. Since according to the inferentialist view of conceptual contents, it is these implicitly recognized material inferential relations that conceptual contents consist in, the conditional permits such contents to be explicitly expressed. If there is a disagreement about the goodness of an inference, it is possible to say what the dispute is about, and offer reasons one way or the other. The conditional is the paradigm of a locution that permits one to make inferential commitments explicit as the contents of judgments. In a similar fashion, introducing negation makes it possible to express explicitly material incompatibilities of sentences, which also contribute to their content. The picture is accordingly one whereby first, formal validity of inferences is defined in terms of materially correct inferences and some privileged vocabulary; second, that privileged vocabulary is identified as logical vocabulary; and third, what it is for something to be a bit of logical vocabulary is explained in terms of its semantically expressive role.

Frege is not as explicit about the role of materially correct inferences as Sellars is, but his commitment to the notion is clear from the relation between two of the views that have been extracted from the *Begriffsschrift*: expressivism about logic and inferentialism about content. Expressivism about logic means that Frege treats logical vocabulary as having a distinctive expressive role--making explicit the inferences that are implicit in the conceptual contents of nonlogical concepts. Inferentialism about those conceptual contents is taking them to be identified and individuated by their inferential roles. Together these views require that it be coherent to talk about inference prior to the introduction of specifically logical vocabulary, and so prior to the identification of any inferences as good in virtue of their form. In the context of an inferential understanding of conceptual contents, an expressivist approach presupposes a notion of nonlogical inference, the inferences in virtue of which concepts have nonlogical content. Thus the young Frege envisages a field of material inferences that confer conceptual content on sentences caught up in them. So although Frege does not offer an explanation of the concept, in the Begriffsschrift his expressive, explicitating project commits him to something playing the role Sellars later picks out by the phrase "material inference".

VIII: Dummett's Model, and Gentzen

So far three themes have been introduced:

--that conceptual content is to be understood in terms of role in reasoning rather than exclusively in terms of representation,

--that the capacity for such reasoning is not to be identified exclusively with mastery of a logical calculus, and

--that besides theoretical and practical reasoning using contents constituted by their role in material inferences, there is a kind of expressive rationality that consists in making implicit content-conferring inferential commitments explicit as the contents of assertible commitments. In this way, the material inferential practices, which govern and make possible the game of giving and asking for reasons, are brought into that game, and so into consciousness, as explicit topics of discussion and justification.

These three themes, to be found in the early works of both Frege and Sellars, provide the beginnings of the structure within which modern inferentialism develops. These ideas can be made more definite by considering a general model of conceptual contents as inferential roles that has been recommended by Dummett. According to that model, the use of any linguistic expression or concept has two aspects: the *circumstances* under which it is correctly applied, uttered, or used, and the appropriate *consequences* of its application, utterance, or use. Though Dummett does not make this point, this model can be connected to inferentialism via the principle that the content to which one is committed by using the concept or expression may be represented by the inference one implicitly endorses by such use, the inference, namely, from the circumstances of appropriate employment to the appropriate consequences of such employment.

The original source for the model lies in a treatment of the grammatical category of sentential connectives. Dummett's two-aspect model is a generalization of a standard way of specifying the inferential roles of logical connectives, due ultimately to Gentzen. Gentzen famously defined connectives by specifying introduction rules, or inferentially sufficient conditions for the employment of the connective, and elimination rules, or inferentially necessary consequences of the employment of the connective. So, to define the inferential role of an expression '&' of Boolean conjunction, one specifies that anyone who is committed to p, and committed to q, is thereby to count also as committed to p&q, and that anyone who is committed to p&q is thereby committed both to p and to q. The first schema specifies, by means of

expressions that do not contain the connective, the circumstances under which one is committed to claims expressed by sentences that do contain (as principle connective) the connective whose inferential role is being defined, that is, the sets of premises that entail them. The second schema specifies, by means of expressions that do not contain the connective, the consequences of being committed to claims expressed by sentences that do contain (as principle connective) the connective whose inferential role is being defined, that is, the sets of consequences that they entail.

IX: Circumstances and Consequences for Sentences

Dummett makes a remarkable contribution to inferentialist approaches to conceptual content by showing how this model can be generalized from logical connectives to provide a uniform treatment of the meanings of expressions of other grammatical categories, in particular sentences, predicates and common nouns, and singular terms. The application to the propositional contents expressed by whole sentences is straightforward. What corresponds to an *introduction* rule for a propositional content is the set of *sufficient* conditions for asserting it, and what corresponds to an *elimination* rule is the set of *necessary* consequences of asserting it, that is, what follows from doing so. Dummett says:

14] Learning to use a statement of a given form involves, then, learning two things: the conditions under which one is justified in making the statement; and what constitutes acceptance of it, i.e., the consequences of accepting it.[30]

Dummett presents his model as specifying two fundamental features of the *use* of linguistic expressions, an idea I'll return to below. In what follows here, though, I'll be applying it in the context of the previous ideas to bring into relief the implicit material inferential *content* a concept or expression acquires in virtue of being used in the ways specified by these two 'aspects'. The link between pragmatic significance and inferential content is supplied by the fact that asserting a sentence is implicitly undertaking a commitment to the correctness of the material inference from its circumstances to its consequences of application.

Understanding or grasping a propositional content is here presented not as the turning on of a Cartesian light, but as practical mastery of a certain kind of inferentially articulated doing: responding differentially according to the circumstances of proper application of a concept, and distinguishing the proper inferential consequences of such application. This is not an all-or-none affair; the metallurgist understands the concept tellurium better than I do, for training has made her master of the inferential intricacies of its employment in a way that I can only crudely

[30] Dummett, *FPL* p. 453.

approximate. Thinking clearly is on this inferentialist rendering a matter of knowing what one is committing oneself to by a certain claim, and what would entitle one to that commitment. Writing clearly is providing enough clues for a reader to infer what one intends to be committed to by each claim, and what one takes it would entitle one to that commitment. Failure to grasp either of these components is failure to grasp the inferential commitment use of the concept involves, and so failure to grasp its conceptual content.

Failure to think about both the circumstances and consequences of appliction leads to semantic theories that are literally one-sided. Verificationists, assertibilists, and reliabilists make the mistake of treating the *first* aspect as exhausting content. Understanding or grasping a content is taken to consist in practically mastering the circumstances under which one becomes entitled or committed to endorse a claim, quite apart from any grasp of what one becomes entitled or committed to by such endorsement. But this cannot be right. For claims can have the same circumstances of application and different consequences of application, as for instance 'I foresee that I will write a book about Hegel' and 'I will write a book about Hegel' do. We can at least regiment a use of ‘foresee’ that makes the former sentence have just the same assertibility conditions as the latter. But substituting the one for the other turns the very safe conditional “If I will write a book about Hegel, then I will write a book about Hegel,” into the risky “If I *foresee* that I will write a book about Hegel, then I will write a book about Hegel.” The possibility that I might be hit by a bus does not affect the assessment of the inference codified by the first conditional, but is quite relevant to the assessment of the second inference.

And the point of the discussion of Sellars' application of inferentialist ideas to the understanding of noninferential reports, at the beginning of this essay, was that parrots and photocells and so on might reliably discriminate the circumstances in which the concept 'red' should be applied, without thereby grasping that concept, precisely in the case where they have no mastery of the consequences of such application—when they can't tell that it follows from something being red that it is colored, that it is not a prime number, and so on. You do not convey to me the content of the concept 'gleeb' by supplying me with an infallible gleebness tester, which lights up when and only when exposed to gleeb things. I would in that case know what things were gleeb, without knowing what I was saying about them when I called them that, what I had found out

about them or committed myself to. Dummett offers two examples of philosophically important concepts where it is useful to be reminded of this point:

15] An account, however accurate, of the conditions under which some predicate is rightly applied may thus miss important intuitive features of its meaning; in particular, it may leave out what we take to be the point of our use of the predicate. A philosophical account of the notion of truth can thus not necessarily be attained by a definition of the predicate 'true', even if one is possible, since such a definition may be correct only in the sense that it specifies correctly the application of the predicate, while leaving the connections between this predicate and other notions quite obscure.[31]

Even more clearly:

16] A good example would be the word 'valid' as applied to various forms of argument. We might reckon the syntactic characterization of validity as giving the criterion for applying the predicate 'valid' to an argument, and the semantic characterization of validity of giving the consequences of such an application. ...if he is taught in a very unimaginative way, he may see the classification of arguments into valid and invalid ones as resembling the classification of poems into sonnets and non-sonnets, and so fail to grasp that the fact that an argument is valid provides any grounds for accepting the conclusion if one accepts the premises. We should naturally say that he had missed the point of the distinction.[32]

Pragmatists of the classical sort, on the other hand, make the converse mistake of identifying propositional contents exclusively with the *consequences* of endorsing a claim, looking downstream to the claim's role as a premise in practical reasoning and ignoring its proper antecedents upstream. [For present purposes, that the emphasis is on *practical* consequences doesn't matter.] Yet one can know what follows from the claim that someone is responsible for a particular action, that an action is immoral or sinful, that a remark is true or in bad taste, without for that reason counting as understanding the claims involved, if one has no idea when it is

[31] Dummett, *FPL* p. 455.
[32] Dummett, *FPL* pp. 453-4.

appropriate to make those claims or apply those concepts. Being classified as AWOL does have the consequence that one is liable to be arrested, but the specific circumstances under which one acquires that liability are equally essential to the concept.

X: 'Derivation', Prior, Belnap, and Conservativeness

Of course, such one-sided theories don't simply ignore the aspects of content they don't treat as central. Dummett says:

17] ...most philosophical observations about meaning embody a claim to perceive... a simple pattern: the meaning of a sentence consists in the conditions for its truth and falsity, or in the method of its verification, or in the practical consequences of accepting it. Such dicta cannot be taken to be so naive as to involve overlooking the fact that there are many other features of the use of a sentence than the one singled out as being that in which its meaning consists: rather, the hope is that we shall be able to give an account of the connection that exists between the different aspects of meaning. One particular aspect will be taken as central, as constitutive of the meaning of any given sentence...; all other features of the use of the sentence will then be explained by a uniform account of their derivation from that feature taken as central.[33]

I think this is a very helpful way to think about the structure of theories of meaning in general, but two observations should be made. First, the principle that the task of a theory of meaning is to explain the use of expressions to which meanings are attributed does not mandate identifying meaning with an aspect of use. Perhaps meanings are to use as theoretical entities are to the observable ones whose antics they are postulated to explain. We need not follow Dummett in his semantic instrumentalism. Second, one might deny that there are meanings in this sense: that is deny that all the features of the use of an expression can be derived in a uniform way from anything we know about it. Dummett suggests that this is the view of the later Wittgenstein. One who takes language to be a motley in this sense will deny that there are such a things as meanings to be the objects of a theory (without, of course, denying that expressions are meaningful). Keeping these caveats in mind, we will find that pursuing this notion of *derivation* provides a helpful perspective on the idea of conceptual contents articulated according to material inferences, and on the role of explicit inference licenses such as conditional statements in expressing and elucidating such inferences, and so such contents.

[33] Dummett, *FPL* pp. 456-7.

For the special case of defining the inferential roles of logical connectives by pairs of sets of rules for their introduction and for their elimination, which motivates Dummett's broader model, there is a special condition it is appropriate to impose on the relation between the two sorts of rules.

18] In the case of a logical constant, we may regard the introduction rules governing it as giving conditions for the assertion of a statement of which it is the main operator, and the elimination rules as giving the consequences of such a statement: the demand for harmony between them is then expressible as the requirement that the addition of the constant to a language produces a conservative extension of that language.[34]

Recognition of the appropriateness of such a requirement arises from consideration of connectives with 'inconsistent' contents. As Prior pointed out, if we define a connective, which after Belnap we may call 'tonk', as having the introduction rule proper to disjunction and the elimination rule proper to conjunction, then the first rule licenses the transition from p to p tonk q, for arbitrary q, and the second licenses the transition from p tonk q to q, and we have what he called a "runabout inference ticket" permitting any arbitrary inference. Prior thought that this possibility shows the bankruptcy of Gentzen-style definitions of inferential roles. Belnap shows rather that when logical vocabulary is being introduced, one must constrain such definitions by the condition that the rule not license any inferences involving only old vocabulary that were not already licensed before the logical vocabulary was introduced, that is, that the new rules provide an inferentially conservative extension of the original field of inferences. Such a constraint is necessary and sufficient to keep from getting into trouble with Gentzen-style definitions. But the expressive account of what distinguishes logical vocabulary shows us a deep *reason* for this demand; it is needed not only to avoid horrible consequences but because otherwise logical vocabulary cannot perform its expressive function. Unless the introduction and elimination rules are inferentially conservative, the introduction of the new vocabulary licenses new material inferences, and so alters the contents associated with the old vocabulary. So if logical vocabulary is to play its distinctive expressive role of making explicit the original material inferences, and so conceptual contents expressed by the old vocabulary, it must be a criterion of

[34] Dummett, *FPL* p. 454.

adequacy for introducing logical vocabulary that no new inferences involving only the old vocabulary be made appropriate thereby.

XI: 'Boche' and the Elucidation of Inferential Commitments

The problem of what Dummett calls a lack of 'harmony' between the circumstances and the consequences of application of a concept may arise for concepts with material contents, however. Seeing how it does provides further help in understanding the notion of expressive rationality, and the way in which the explicitating role of logical vocabulary contributes to the clarification of concepts. For conceptual change can be:

19] ...motivated by the desire to attain or preserve a harmony between the two aspects of an expression's meaning. A simple case would be that of a pejorative term, e.g. 'Boche'. The conditions for applying the term to someone is that he is of German nationality; the consequences of its application are that he is barbarous and more prone to cruelty than other Europeans. We should envisage the connections in both directions as sufficiently tight as to be involved in the very meaning of the word: neither could be severed without altering its meaning. Someone who rejects the word does so because he does not want to permit a transition from the grounds for applying the term to the consequences of doing so. The addition of the term 'Boche' to a language which did not previously contain it would produce a non-conservative extension, i.e. one in which certain other statements which did not contain the term were inferable from other statements not containing it which were not previously inferable...[35]

This crucial passage makes a number of points that are worth untangling. First of all, it shows how concepts can be criticized on the basis of substantive beliefs. If one does not believe that the inference from German nationality to cruelty is a good one, one must eschew the concept or expression "Boche". For one cannot deny that there are any Boche--that is just denying that anyone is German, which is patently false. One cannot admit that there are Boche and deny that they are cruel--that is just attempting to take back with one claim what one has committed oneself to with another. One can only refuse to employ the concept, on the grounds that it embodies an inference one does not endorse.

[35] Dummett, *FPL* p. 454.

The prosecutor at Oscar Wilde's trial at one point read out some of the more hair-raising passages from "The Importance of Being Earnest" and said "I put it to you, Mr. Wilde, that this is *blasphemy*. Is it or is it not?" Wilde made exactly the reply he ought to make—indeed, the only one he could make—given the considerations being presented here and the circumstances and consequences of application of the concept in question. He said "Sir, 'blasphemy' is not one of my words."

Highly charged words such as "whore", "faggot", "lady", "Communist", "Republican", have seemed to some a special case because they couple 'descriptive' circumstances of application to 'evaluative' consequences. But this is [are] not the only sort of expression embodying inferences that requires close scrutiny. As I have emphasized, the use of *any* concept or expression involves commitment to an inference from its grounds to its consequences of application. Critical thinkers, or merely fastidious ones, must examine their idioms to be sure that they are prepared to endorse and so defend the appropriateness of the material inferential transitions implicit in the concepts they employ. In Reason's fight against thought debased by prejudice and propaganda, the first rule is that potentially controversial material inferential commitments should be made explicit as claims, exposing them both as vulnerable to reasoned challenge and as in need of reasoned defense. They must not be allowed to remain curled up inside loaded phrases such as "enemy of the people" or "law and order."

It is in this process that formal logical vocabulary such as the conditional plays its explicitating role. It permits the formulation, as explicit claims, of the inferential commitments that otherwise remain implicit and unexamined in the contents of material concepts. Logical locutions make it possible to display the relevant grounds and consequences, and to assert their inferential relation. Formulating as an explicit claim the inferential commitment implicit in the content brings it out into the open as liable to challenges and demands for justification, just as with any assertion. In this way explicit expression plays an elucidating role, functioning to groom and improve our inferential commitments, and so our conceptual contents—a role, in short, in the practices of reflective rationality or "Socratic method".

But if Dummett is suggesting that what is wrong with the concept 'Boche' is that its addition represents a nonconservative extension of the rest of the language, he is mistaken. Its nonconservativeness just shows that it has a substantive content, in that it implicitly involves a material inference that is not already implicit in the contents of other concepts being employed. Outside of logic, this is no bad thing. Conceptual progress in science often consists in

introducing just such novel contents. The concept of temperature was introduced with certain criteria or circumstances of appropriate application, and certain consequences of application. As new ways of measuring temperature are introduced, and new theoretical and practical consequences of temperature measurements adopted, the complex inferential commitment that determines the significance of using the concept of temperature evolves.

The proper question to ask in evaluating the introduction and evolution of a concept is not whether the inference embodied is one that is already endorsed, so that no new content is really involved, but rather whether that inference is one that *ought* to be endorsed. The problem with 'Boche' is not that once we explicitly confront the material inferential commitment that gives them their content, it turns out to be novel, but that it can then be seen to be indefensible and inappropriate--a commitment we cannot become entitled to. We want to be aware of the inferential commitments our concepts involve, to be able to make them explicit, and to be able to justify them. But there are other ways of justifying them than showing that we were already implicitly committed to them, before introducing or altering the concept in question.

XII: Harmony and Material Inference

Even in the cases where it does make sense to identify harmony of circumstances and consequences with inferential conservativeness, the attribution of conservativeness is always relative to a background set of material inferential practices, the ones that are conservatively extended by the vocabulary in question. Conservativeness is a property of the conceptual content only in the context of other contents, not something it has by itself. Thus there can be pairs of logical connectives, either of which is all right by itself, but both of which cannot be included in a consistent system. It is a peculiar ideal of harmony that would be realized by a system of conceptual contents such that the material inferences implicit in every subset of concepts represented a conservative extension of the remaining concepts, in that no inferences involving only the remaining ones are licensed that are not licensed already by the contents associated just with those remaining concepts. Such a system is an idealization, because all of its concepts would already be out in the open; none remaining hidden, to be revealed only by drawing conclusions from premises that have never been conjoined before, following out unexplored lines of reasoning, drawing consequences one was not previously aware one would be entitled or committed to by some set of premises. In short, this would be a case where Socratic reflection, making implicit commitments explicit and examining their consequences and possible justifications, would never motivate one to alter contents or commitments. Such complete transparency of commitment and entitlement is in some sense an ideal projected by the sort of Socratic practice that finds current contents and commitments wanting by confronting them with each other, pointing out inferential features of each of which we were unaware. But as Wittgenstein teaches in general, it should not be assumed that our scheme is like this, or depends upon an underlying set of contents like this, just because we are obliged to remove any particular ways in which we discover it to fall short.

These are reasons to part company with the suggestion, forwarded in the passage above, that inferential conservatism is a necessary condition of a 'harmonious' concept--one that won't 'tonk up' a conceptual scheme. In a footnote, Dummett explicitly denies that conservativeness can in general be treated as a sufficient condition of harmony:

20] This is not to say that the character of the harmony demanded is always easy to explain, or that it can always be accounted for in terms of the notion of a conservative extension. ...the most difficult case is probably the vexed problem of personal identity.[36]

In another place, this remark about personal identity is laid out in more detail:

21] We have reasonably sharp criteria which we apply in ordinary cases for deciding questions of personal identity: and there are also fairly clear consequences attaching to the settlement of such a question one way or the other, namely those relating to ascriptions of responsibility, both moral and legal, to the rights and obligations which a person has... What is much harder is to give an account of the connection between the criteria for the truth of a statement of personal identity and the consequences of accepting it. We can easily imagine people who use different criteria from ours...Precisely what would make the criteria they used criteria for personal identity would lie in their attaching the same consequence, in regard to responsibility, motivation, etc., to their statements of personal identity as we do to ours. If there existed a clear method for deriving, as it were, the consequences of a statement from the criteria for its truth, then the difference between such people and ourselves would have the character of a factual disagreement, and one side would be able to show the other to be wrong. If there were no connection between truth-

[36] Dummett, *FPL* p. 455n.

grounds and consequences, then the disagreement between us would lie merely in a preference for different concepts, and there would be no right or wrong in the matter at all.[37]

Dummett thinks that there is a general problem concerning the way in which the circumstances and consequences of application of expressions or concepts ought to fit together. Some sort of 'harmony' seems to be required between these two aspects of the use. The puzzling thing, he seems to be saying, is that the harmony required cannot happily be assimilated either to compulsion by facts or to the dictates of freely chosen meanings. But the options: matter of fact or relation of ideas, expression of commitment as belief or expression of commitment as meaning are not ones that readers of "Two Dogmas of Empiricism" ought to be tempted to treat as exhaustive.

The notion of a completely factual issue that Dummett appeals to in this passage is one in which the applicability of a concept is settled straightforwardly by the application of other concepts, the concepts that specify the necessary and sufficient conditions that determine the truth conditions of claims involving the original concept. This conception, envisaged by a model of conceptual content as necessary and sufficient conditions, seems to require a conceptual scheme that is ideally transparent in the way mentioned above, in that it is immune to Socratic criticism. For that conception insists that these coincide in that the jointly sufficient conditions already entail the individually necessary ones, so that it is attractive to talk about content as truth conditions, rather than focussing on the substantive inferential commitments that relate the sufficient to the distinct necessary conditions, as recommended here. By contrast to this either/or, in a picture according to which conceptual contents are conferred on expressions by their being caught up in a structure of inferentially articulated commitments and entitlements, material inferential commitments are a necessary part of any package of practices that includes material doxastic commitments.

The circumstances and consequences of application of a nonlogical concept may stand in a substantive material-inferential relation. To ask what sort of 'harmony' they should exhibit is to ask what material inferences we ought to endorse, and so what conceptual contents we ought to employ. This is not the sort of a question to which we ought to expect or welcome a general or wholesale answer. Grooming our concepts and material-inferential commitments in the light of our assertional commitments, including those we find ourselves with noninferentially through observation, and the latter in the light of the former, is a messy, retail business.

Dummett thinks that a theory of meaning should take the form of an account of the nature of the 'harmony' that ought to obtain between the circumstances and the consequences of application of the concepts we ought to employ. Moving up a level now to apply these considerations about the relations of circumstances to consequences of appliation to the contents of the concepts employed in the metalanguage in which we couch a semantic theory, the point I want to make is that we should not expect a theory of that sort to take the form of a specification

[37] Dummett, *FPL* p. 358.

of necessary and sufficient conditions for the circumstances and consequences of application of a concept to be harmonious. For that presupposes that the circumstances and consequences of application of the concept of harmony do not themselves stand in a substantive material inferential relation. On the contrary, insofar as the idea of a theory of semantic or inferential harmony makes sense at all, it must take the form of an investigation of the ongoing elucidative process, of the 'Socratic method' of discovering and repairing discordant concepts, which alone gives the notion of harmony any content. It is given content only by the process of harmonizing commitments, from which it is abstracted. In Sellars' characterization of expressive rationality, modal claims are assigned the expressive role of inference licenses, which make explicit a commitment that is implicit in the use of conceptual contents antecedently in play. Rules of this sort assert an authority over future practice, and answer for their entitlement both to the prior practice being codified and to concomitant inferential and doxastic commitments. In this way they may be likened to the principles formulated by judges at common law, intended both to codify prior practice, as represented by precedent, expressing explicitly as a rule what was implicit therein, and to have regulative authority for subsequent practice. The expressive task of making material inferential commitments explicit plays an essential role in the reflectively rational Socratic practice of harmonizing our commitments. For a commitment to become explicit is for it to be thrown into the game of giving and asking for reasons as something whose justification, in terms of other commitments and entitlements, is liable to question. Any theory of the sort of inferential harmony of commitments we are aiming at by engaging in this reflective, rational process must derive its credentials from its expressive adequacy to that practice, before it should be accorded any authority over it.

XIII: From Semantics To Pragmatics

In the first part of this essay I introduced three related ideas:

--the *inferential* understanding of conceptual content,

--the idea of *materially* good inferences, and

--the idea of *expressive* rationality.

These contrast, respectively, with:

--an understanding of content exclusively according to the model of the *representation* of states of affairs, [I think I've managed to say rather a lot about conceptual content in this essay, without talking at all about what is represented by such contents.]

--an understanding of the goodness of inference exclusively on the model of *formal* validity, and

--an understanding of rationality exclusively on the model of *instrumental* or means-end reasoning.

In the second part of the essay, these ideas were considered in relation to the representation of inferential role suggested by Dummett, in terms of the circumstances of appropriate application of an expression or concept and the appropriate consequences of such application. It is in the context of these ideas that I have sought to present an *expressive* view of the role of logic, and its relation to the practices constitutive of rationality. That view holds out the hope of recovering for the study of *logic* a direct significance for projects that have been at the core of *philosophy* since its Socratic inception.

END of Lecture 3

On the Way to a Pragmatist Theory of the Categories[38]

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). [39]

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.

Rorty called those impressed by the scientific naturalism epitomized in the *scientia mensura* "right wing Sellarsians," and those impressed by the normative nonnaturalism about

---

[38] This lecture is drawn from the Introduction to *From Empiricism to Expressivism: Brandom Reads Sellars* [Harvard University Press, 2015].

[39] 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).

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.[40]) 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.[41]

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.)

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

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

[41] 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.

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.[42] The neo-Kantian tradition comprising Lewis, Carnap, and Sellars can be thought of as an undercurrent, somewhat occluded from view by the empiricist surface.

2. Categories in Kant

Many Kantian themes run through Sellars's philosophy. My Sellars 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.[43] Among the vocabularies that play the second sort of role, Sellars includes

[42] 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].

[43] 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.

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."

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*.[44]

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.[45]

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.

Kant, however, looks at Newton's formulation of the best empirical understanding of his day and sees that the newly introduced concepts of <u>force</u> and <u>mass</u> 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 <u>force</u> and <u>mass</u>, but are using some at least slightly different concepts. (Galileo's geometrical version of the (late

[44] CDCM §79.

[45] 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.

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 nonetheless uncovered a crucial *semantic* difference between the expressive roles of different kinds of concepts. Once his attention had been directed to them, Kant 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.[46] 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.)

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*.[47] 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).

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

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

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

broad sense *metalinguistic*. In pursuing this line he follows Rudolph Carnap, who in addition to 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. 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.[48] 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.

> 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

[48] "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]

> label. The descriptive and explanatory resources of language advance hand in hand.[49]

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.

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

[49] CDCM §108.

'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".[50]

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.[51]

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.)

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

[50] Sellars, "Inference and Meaning" *PPPW* pp. 265/317, reprinted in *In the Space of Reasons.*
[51] Sellars, "Inference and Meaning" *PPPW* pp. 284/336.

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',[52]

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.[53] 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.

---

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

[53] 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 dog.

Fido is a brute.

Fido is an animal.

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. [54]

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.[55]

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.[56]

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.[57]

---

[54] "Towards a Theory of the Categories" (TTC) §10-11.
[55] TTC §16.
[56] TTC §22.
[57] TTC §23.

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.

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.

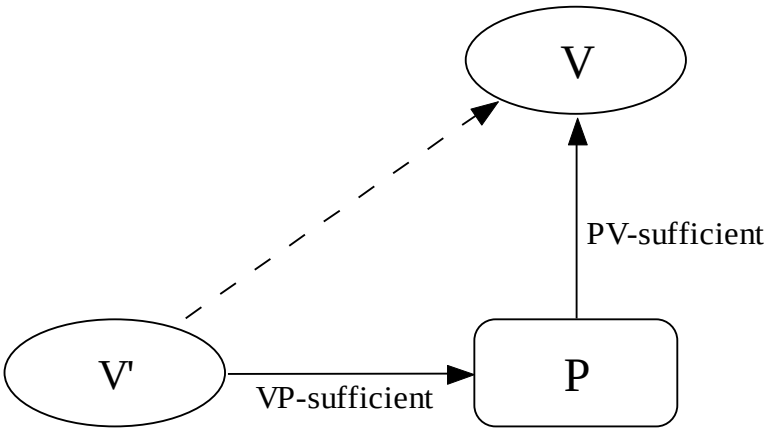
- *Fourth*, the features of vocabulary(concept)-use they explicate are *universal*: they are features of any and every autonomous discursive practice.

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:[58]

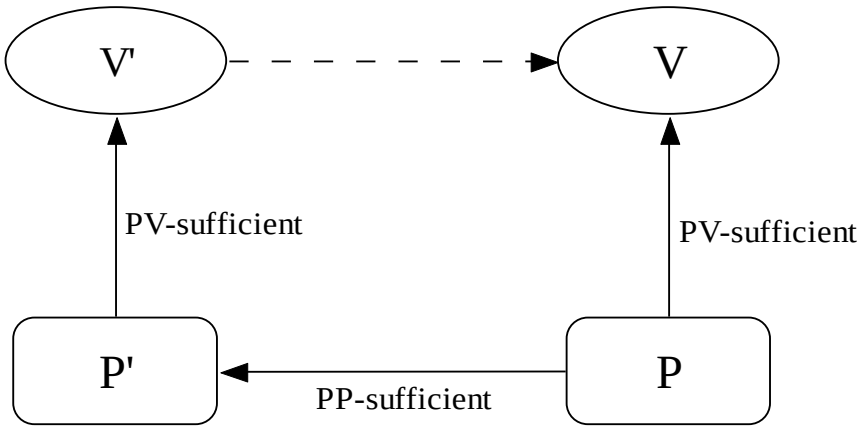
[58] I introduce, develop, and apply these "meaning-use diagrams" in *Between Saying and Doing: Towards an Analytic Pragmatism* [Oxford University Press, 2008].



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.

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

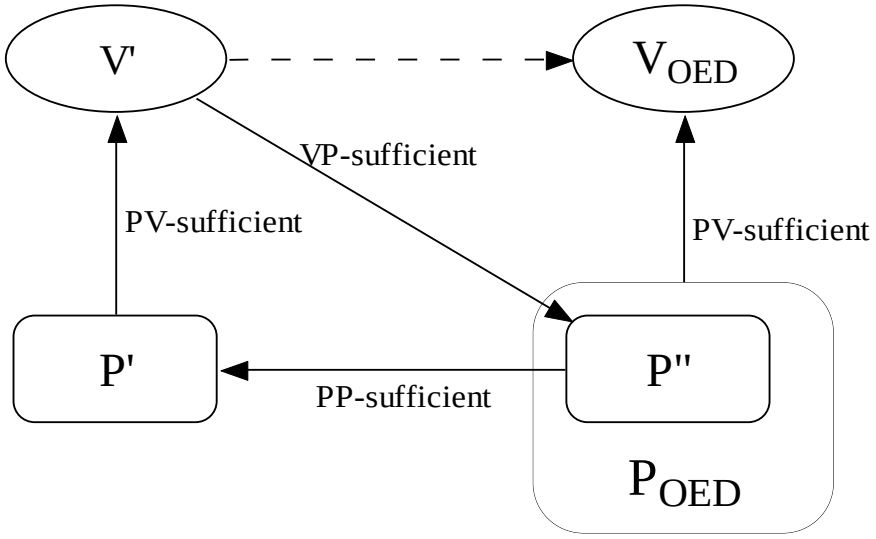
from the old vocabulary by using the new vocabulary. On the side of circumstances of application (assertibility conditions), one must acknowledge commitment to the conditional p→q just in case one takes the inference from p to q to be a materially good one. And on the side of consequences of application, 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 are 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*.[59] The feature of the LX model that corresponds to the a prioricity of

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

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 *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.

END of Lecture 4

Modal Expressivism and Modal Realism: Together Again[60]

I. A Modal Expressivism

1. Kant saw that in addition to concepts whose principal use is to make it possible for us to describe how things are, there are concepts that make explicit features of the metaconceptual *framework* that makes such description possible. An important class of the framework-explicating concepts (arguably the one that motivated this entire line of thought) comprises *alethic modal* concepts, such as necessity and possibility. These express lawful relations between ground-level descriptive concepts, and mark the special status of Newton's laws, their lawfulness, by contrast to the status of merely contingent matters of fact. But it is not only in understanding the use of technical scientific concepts that the modal concepts find application. The use of ordinary empirical descriptive concepts such as gold, and cat, and house, no less than the Newtonian concepts of mass, force, and acceleration, is essentially, and not just accidentally, articulated by the modality these modal concepts express.

It is because he believes all this that Kant calls modal concepts (among others) 'pure' concepts: categories. Pure concepts are a species of *a priori* concepts.[61] The sense in which we can think of them as available *a priori* that I want to focus on comprises three claims. First, what they express are structural features of the framework within which alone it is possible to apply *any* concepts, make *any* judgments, including ordinary empirical descriptive ones. Second, in being able to apply any ground-level empirical concepts, one already knows how to do

[60] The material from which this lecture is drawn was originally published in *From Empiricism to Expressivism: Brandom Reads Sellars* [Harvard University Press, 2015].

61 That is, concepts available *a priori*. I take it that Kant's standard usage of "a priori" is adverbial, though this is not obvious since the Latin phrase is not grammatically marked as it would be in German. *Exactly* what Kant means by the term 'pure' [rein], as it applies generically to reason, knowledge, understanding, principles, concepts, and intuition is a complex and challenging question. There seems to be some terminological drift across the species, and some wavering on how to classify particular examples. (The status of the crucial *a priori* principle that every alteration must have a cause, for instance, is apparently variously characterized at [B3] and [B5].) Being available *a priori* is necessary, but not sufficient [B3].

everything one needs to know how to do in order to apply the categorial concepts. Finally, there are no *particular* empirical descriptive concepts one must be able to apply in order to have implicit mastery of what is expressed by categorial concepts such as the modal ones (though one must have some descriptive concepts or other).

The alethic modality that has this categorial status is something like physical necessitation. It is the modality involved in the "pure principle" that "every alteration must have a cause."

2. A further development of what I want to claim will be retrospectively recognizable as the same line of thought can be found in Frege.[62] His use of Latin letters and his logical sign of generality (used in conjunction with the notation for hypotheticals) express relations between concepts. It has always been an embarrassment for the anachronistic extensional quantificational reading of this notation (due originally to Russell) that Frege says of it, when he first introduces it in the *Begriffsschrift*, that it is the right way to express *causal* relations of necessitation.[63] For it is a commonplace of the later logistical tradition that merely quantificational relations between concepts cannot distinguish between contingent regularities and lawlike, necessary ones. For that, explicit modal operators must be applied to the quantified conditionals.

But Frege deploys his notation so that the relations between concepts expressed by generalized conditionals *already* have modal force. Relations between concepts of the sort logic lets us express have consequences for relations between their extensions, of the sort our quantificational notation expresses, but his generality locutions codify relations we think of as intensional. Fregean logical concepts are indeed second- and higher-order concepts, but more than that, the universality they express is rulish. They are in the first instance principles in

62 The characterization of Frege's Begriffsschrift that follows is one that I had my eyes opened to by Danielle Macbeth's pathbreaking book *Frege's Logic* [Harvard University Press, 2005].

63 "*This is the way in which causal connections are expressed."* [Italics in the original.] *Begriffsschrift* §12 (p. 27 in Jean van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931* [Harvard University Press, 1967]), foreshadowed at §5.

accordance with which to reason, and only derivatively premises from which to reason.[64] In addition to permitting the formulation of purely logical relations among logical concepts, Frege's logical vocabulary permits us to assert necessary connections among empirical concepts that themselves can only be discovered empirically: physically or causally necessary connections. In the Preface to the *Begriffsschrift*, Frege says:

> It seems to me to be easier still to extend the domain of this concept-script to include geometry. We would only have to add a few signs for the intuitive relations that occur there…The transition to the pure theory of motion and then to mechanics and physics could follow at this point. The latter two fields, in which besides rational necessity natural necessity asserts itself, are the first for which we can predict a further development of the notation as knowledge progresses.[65]

The additional signs that such an extension requires do *not* include modal operators. The necessity (whether natural or rational) of the connections between empirical concepts is already contained as part of what is expressed by the *logical* vocabulary, even when it is used to make claims that are not logically, but only empirically true.

3. Nearer to our own time, this line of thought has been further developed and clarified by Wilfrid Sellars. He lucidly compressed his endorsement of the fundamental Kantian idea that modal concepts make explicit something implicit in the use of ordinary empirical descriptive concepts into the title of one of his essays: "Concepts as Involving Laws, and Inconceivable without Them." But he also offers the outline of a more articulated argument for the claim. We can reconstruct it as follows:

1. "It is only because the expressions in terms of which we describe objects… locate these objects in a space of implications, that they describe at all, rather than merely label." [66]

---

64 Following Mill, this is Sellars's way of putting the point, in "Counterfactuals, Dispositions, and the Causal Modalities" Pp. 225-308 of *Minnesota Studies in the Philosophy of Science*, Vol. II, ed. by H. Feigl, M. Scriven, and G. Maxwell, (University of Minnesota Press; Minneapolis, MN: 1957). Henceforth "CDCM."

65 P. 7 in van Heijenoort op.cit.. I have emended the translation slightly, where I have noted the original German terms.

66 "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.] (hereafter CDCM), § 108.

2. It is an essential feature of the inferential relations in which, according to claim (1), descriptive concepts must stand, that they can be appealed to in *explanations* and *justifications* of further descriptions.

3. So: "although describing and explaining (predicting, retrodicting, understanding) are *distinguishable*, they are also, in an important sense, *inseparable*… The descriptive and explanatory resources of language advance hand in hand…."[67]

4. The expressive role distinctive of modal vocabulary is to make explicit these explanatory and justificatory relations.

The two key moves in an argument of this form are, first, an account of the descriptive use of empirical concepts that exhibits as essential their articulation by inferences that can support explanations and justifications, and second, an account of the central function of at least some alethic modal vocabulary as expressing explanatory and justificatory inferential relations among descriptive concepts. The conclusion of the argument is what I call the "Kant-Sellars thesis about modality": in knowing how to use ordinary empirical descriptive vocabulary, one already knows how to do everything one needs to know how to do in order to be able (in principle) to use alethic modal vocabulary.[68] According to this thesis, one cannot be in the semantic predicament that empiricists such as Hume and Quine envisaged: understanding ordinary empirical descriptive vocabulary perfectly well, but having thereby no grip at all on what is expressed by modal vocabulary.

How does Sellars understand the distinction between "merely labeling", on the one hand, and describing, in the sense he then wants to argue "advances hand in hand" with explaining and justifying, on the other hand? Labeling is attaching signs to, or associating them with, items in the nonlinguistic world

It is tempting to think that reliably responding in a distinctive way to some things and not others is a way of *classifying* them as being of some kind, or as having something in common. What more besides dividing things into groups could be required to count as *describing* them as being of different kinds? The difference between classifying in the sense of labeling and

67 CDCM § 108.

68 I discuss this claim at greater length in Chapter Four of *Between Saying and Doing: Towards an Analytic Pragmatism* (Oxford: Oxford University Press, 2008).

describing emerges when we ask what the things grouped together by their elicitation of a common response are supposed to be described *as*. If the dog reliably barks at some things, and not others (cats, dogs, and squirrels, but not horses, men but not women, motorcycles but not cars, helicopters but not airplanes, church bells but not the neighbor's stereo, and so on) it is *grouping* things, sorting them into two classes. But there need be nothing it is *describing* them *as*. When the metal strip expands in some environments and contracts in others, it is not yet *describing* them as warm or cold.

Sellars's idea is that what one is describing something *as* is a matter of what *follows* from the classification—what *consequences* falling in one group or another has. It is insofar as being grouped one way rather than another can serve as a *premise* in an *inference* that the grouping is intelligible as a *description* and not merely a label. Reliably differentially elicited responses are intelligible as observation reports, as empirical descriptions, just insofar as they are available to *justify* further claims. It is essential, and not just accidental, to descriptive predicates that they can be used to make claims, which would be expressed by declarative sentences. And it is essential, and not accidental to those claimings that they can serve as *reasons* for further claims.

Sellars sees modal locutions as tools used in the enterprise of

> …making explicit the rules we have adopted for thought and action…I shall be interpreting our judgments to the effect that A causally necessitates B as the expression of a rule governing our use of the terms 'A' and 'B'.[69]

The rules they express are rules of *inference*. Modal expressions are inference licenses or inference "tickets," in Ryle's terminology.[70] These are what Sellars calls "material," that is, non-logical inferences. In fact, what these modal locutions make explicit, according to Sellars, are just the implications, situation in a space of which is what distinguishes descriptive concepts from mere labels. Inferences such as that from "Pittsburgh is to the West of New York, so New York is to the East of Pittsburgh," articulate the content of the descriptive concepts West and East.

69 Sellars, "Language, Rules, and Behavior" footnote 2 to p. 136/296 in *PPPW*.
70 Gilbert Ryle, " 'If', 'So', and 'Because' ", pp. 302-318 in Black, Max (ed.) *Philosophical Analysis* [Prentice Hall, 1950]. Sellars does not discuss whether "A causally necessitates B" should be understood as expressing a committive, or merely a permissive inference.

Further, it is the inferential commitments acknowledging such material implicational relations that are appealed to in explanation and justification.

> To make first hand use of these [modal] expressions is to be about the business of explaining a state of affairs, or justifying an assertion.[71]

That is, what one is *doing* in *using* modal expressions ("As are necessarily Bs") is endorsing an inference (from anything's being A to its being B) that can be appealed to in justifying one description on the basis of another, or explaining the applicability of one description by the appealing to the applicability of another: "The raspberries are red *because* they are ripe." This is why the expressive resources of description, on the one hand, and justification and explanation, on the other hand, "advance hand in hand," as Sellars says.

This constellation of claims to which Sellars aspires to entitle himself articulates what he makes of the tradition of thinking about modality that Kant initiates and Frege develops in an inferentialist key. It is a story that construes (at least one kind of) modal vocabulary as distinguished by the role it plays in expressing explicitly essential aspects that it makes visible as implicit already in the use of ordinary empirical descriptive vocabulary. Having a ("first hand") use in explicating the framework within which vocabulary use can have the significance of describing—a framework we come to see as necessarily a unified package comprising not only description, but justification and explanation, a framework articulated by subjunctively robust inferential relations among descriptive concepts—sets modal vocabulary off from the descriptive vocabulary, precisely in virtue of the distinctive expressive role it plays with respect to the use of such descriptive vocabulary. This, then, is Sellars's modal expressivism.

4. It is, it should be acknowledged, largely programmatic. Turning the program into a full-blooded account of the use of modal vocabulary would require satisfactory responses to a number of challenges. I remarked above that Sellars's approach focuses on modally qualified conditionals. So, at a minimum, we would need to understand how it might be developed or extended to deal with other uses of modal operators.[72]

---

71 CDCM § 80.

72 Semantic inferentialists think that the use of *any* concept involves commitment to the propriety of all the inferences from the circumstances of appropriate application to the appropriate consequences of application of that

Another challenge to working out Sellars's version of modal expressivism concerns the extent to which, and the sense in which, it should be understood as taking the expressive role characteristic of modal vocabulary to be a *metalinguistic* one. On the one hand, when Sellars says he wants to understand a paradigmatic kind of modal judgment as "the expression of a rule governing our use of the terms 'A' and 'B'," this sounds straightforwardly metalinguistic in a classical sense. On the other hand, it cannot be right to say that modal claims should be understood as covertly made in a metalanguage whose mastery requires mastery of terms that *refer to* terms (here, descriptive ones) in an object language—which is the classical Tarski-Carnap sense. For someone (perhaps a monolingual German) could claim, believe, or judge that A causally necessitates B without ever having heard of the English *expressions* that 'A' and 'B' stand for in the example. Further, the claim could be true even if there had never been such expressions, because there had never been any language users. (There would still have been laws of nature, even if there had never been language.) So is the view he is after a metalinguistic expressivism, or not? In light of the considerations just mentioned, Sellars's characteristically nuanced-but-unhelpful assessment is this:

> Shall we say that modal expressions are metalinguistic? Neither a simple 'yes' nor a simple 'no' will do.[73]

He wants to say that while modal statements are not metalinguistic in a narrow sense, there is a wider sense in which they are.

> It is sometimes thought that modal statements do not describe states of affairs in the world, because they are *really* metalinguistic. This won't do at all if it is meant that instead of describing states of affairs in the world, they describe linguistic habits. It is more plausible if it is meant that statements involving modal terms have the force of *prescriptive* statements about the use of certain expressions in the object language. Yet there is more than one way of to '*have*

concept. (Cf. Chapter One of *Articulating Reasons* [Cambridge: Harvard University Press, 1997]. So in that context, a strategy for addressing this challenge might not be far to seek.

73 CDCM §82.

> *the force of*" a statement, and failure to distinguish between them may snowball into a serious confusion as wider implications are drawn.[74]

I think Sellars never really figures out how to work out the line of thought he suggests here. After 1959 he never repudiates the views he sketched in "Counterfactuals, Dispositions, and the Causal Modalities," and seems to continue to endorse them. But he never revisits the topic substantially—never says how he thinks one might go on to fill in the expressivist idea he had gestured at there. Doing that is, in effect, left as an exercise to the reader.

5. Sellars is working with Kant's idea that the expressive role distinctive of alethic modal vocabulary is to make explicit something that is implicit already in the use of ordinary empirical descriptive vocabulary. He picks up Frege's hint that what matters is the specifically *inferential* articulation essential to the conceptual contentfulness of descriptive vocabulary. He develops those thoughts by adding the ideas that that expressive role is in some broad but noncanonical sense metalinguistic—a matter of the role such vocabulary plays in endorsing rules of inference governing descriptive vocabulary. And equally importantly, he focuses our attention on the *pragmatic* dimension of that expressive role. That is, he counsels us to look to what we are *doing* when we endorse a modal claim.

I want to make a couple of suggestions for how one might move forward with what Sellars made of Kant's thought about how the expressive role characteristic of alethic modal vocabulary is related to that of ordinary empirical descriptive vocabulary. One lesson I think we can learn from Sellars's difficulties is that the notion of being 'metalinguistic' or ("about language") is too crude an expressive tool, too undifferentiated a concept, to be helpful in this context. There are, as Sellars intimates, *many* ways in which the use of one vocabulary can depend on that of another, besides any terms of the one vocabulary *referring* to those of the other. Putting together Sellars's *metalinguistic* idea with his *pragmatic* idea, we could consider the possibility that the place to begin thinking about the expressive role of modal vocabulary is with what in *Between Saying and Doing* I call a "pragmatic metavocabulary." This concept

74 CDCM §81.

takes its place alongside that of a syntactic metavocabulary, which enables one to talk about linguistic expressions themselves (both what Sellars calls "sign designs" and grammatical categories), and a semantic metavocabulary, which enables one to talk about what linguistic expressions refer to or what descriptive concepts let one say. A *pragmatic* metavocabulary enables one to talk about what one is *doing* in *using* linguistic expressions, the speech acts one is performing, the pragmatic force one is investing them with or exercising, the commitments one is undertaking by making claims, the norms that govern linguistic performances, and so on. Sellars's model is that modal vocabulary says something that would be said more explicitly in a *semantic* metavocabulary. But by the time his commentary has taken back everything that it turns out needs to be taken back, not much is left of that model. What seems right about the commentary, however, is Sellars's observations about what one is *doing* in "making first-hand use" of modal vocabulary: endorsing inferences. Insofar as there is anything to that idea, the more natural strategy would seem to be to take one's model from *pragmatic* metavocabularies. After all, Sellars ends up saying nothing at all about what one *says* in making first-hand use of modal vocabulary. Properly understood, I think, his is not a *semantic* expressivism about alethic modal vocabulary, but a kind of *pragmatic* expressivism about it.

As a first try at expressing the thought that would result from transposition from a semantic into a pragmatic key, we might try this: In making first-hand use of (the relevant kind of) alethic modal vocabulary one is *doing* something distinctive that could be specified explicitly in the right kind of pragmatic metavocabulary, namely endorsing a class of inferences. The pragmatic metavocabulary enables one to *say* what modal vocabulary enables one to *do*. Such a claim does not in itself involve any commitment concerning the relations between the *content* of talk about endorsing inferences and talk about necessity and possibility, never mind commitment to their equivalence.

My second suggestion for developing Sellars's modal expressivism is that what is special about (a certain kind of) modal vocabulary is that it stands in a special relation to descriptive vocabulary—a relation that invited its characterization as 'metalinguistic' (with respect to that descriptive vocabulary) in the first place. This relation is that anyone who knows how to use ordinary empirical descriptive vocabulary (e.g. 'red', 'square', 'moving', 'alive', 'electron')

already knows how to do everything she needs to know how to do to deploy modal vocabulary. In this sense, modal vocabulary makes explicit (in the form of a new kind of claimable content) something that is implicit already in the *use* of descriptive vocabulary. Not all vocabularies stand in this relation to some other kind of vocabulary. In particular, there is in general nothing that ordinary empirical descriptive (OED) vocabulary stands to in this expressive relation.

What aspect of inference is it that modal vocabulary is supposed to express? My third suggestion for developing the Kant-Sellars approach to modality is an answer to this question. The key fact to appreciate, I think, is that outside of logic and mathematics, in ordinary language and the special sciences, material inference is massively *nonmonotonic*. That is, the fact that the inference from *p* to *q* is a materially good one in some situation does not mean that the inference from *p* and *r* to *q* must also be a good one, in the same situation. If I strike this dry, well-made match, it will light—but not if in addition all the oxygen is removed from the room, or a sufficiently strong magnetic field is applied, or…. If I let loose of the leash, the dog will chase the cat—but not if either one is struck by lightning, a bear suddenly blocks the way, or…. This phenomenon is ubiquitous and unavoidable, even in less informal contexts: differential medical diagnosis, the application of common or case law, or philosophical argumentation. One cannot secure material inferences from all possible defeasors by explicitly building their denial into the premises, for the class of defeasors is in general open-ended and not antecedently surveyable. Nor can one achieve the same effect wholesale by the use of *ceteris paribus* clauses. As I have argued elsewhere, the expressive role of such clauses is explicitly to acknowledge the non-monotonicity, hence defeasibility of the qualified inference, not magically to remove it. [75] (The technical term for a Latin phrase whose application can do *that* is 'spell').

The defeasibility or nonmonotonicity of the material inferences essential to the conceptual contentfulness of descriptive vocabulary means that the use of such vocabulary requires not only making a distinction (however fallibly) between those inferences one endorses and those one does not, but also (as part of that capacity, and also fallibly) between the collateral premises or auxiliary hypotheses whose additions one takes it would, and those that would not,

75 In Chapter Two of *Articulating Reasons: An Introduction to Inferentialism* [Cambridge: Harvard University Press, 2000].

infirm the inference, in the sense that the conclusion would no longer follow. That is, in order to use OED vocabulary, one must associate *some* range of subjunctive and counterfactual robustness with the material inferences that (at least partially) articulate the contents of the descriptive concepts. So, for instance, I might endorse the inference that would be made explicit in a conditional by "If I release my grip on the book, then it will fall to the floor." But for the attribution of such an inferential commitment to me to be sustainable, I must make some distinction between collateral circumstances that would defeat the inference (a table is moved under it, someone else catches it, it dissolves in a puff of smoke, it is snatched up by a passing hawk…) and those that would not (it is Tuesday, it is slightly cooler today than it was yesterday, my car has been moved slightly further away…). Of course I might be wrong about whether any of these particular auxiliary hypotheses actually would or would not defeat the inference to the conclusion. But if I make no distinction of this sort at all I should be convicted of not understanding the concepts (book, falling) that I am attempting to apply.

On this account, subjunctive robustness is the generality or "openness" Ryle found in the inferences made explicit by conditionals, and which is made explicit by modal vocabulary, including the subjunctive mood. It involves a kind of quantification over auxiliary hypotheses that would not, according to the modal claim, infirm the inference or its conclusion.[76] The kind of generalization implicit in the use of subjunctive or modal vocabulary is what is invoked in *explanation*, which exhibits some conclusion as the resulting from an inference that is good as an instance of a *kind*, or in virtue of a *pattern* of good inferences. It is because the use of descriptive vocabulary requires commitment to inferences with some range of subjunctive robustness that, as I earlier quoted Sellars as saying:

> Although describing and explaining (predicting, retrodicting, understanding) are *distinguishable*, they are also, in an important sense, *inseparable*… The descriptive and explanatory resources of language advance hand in hand…."[77]

The expressive job characteristic of modal vocabulary is to make explicit this implicit dimension of the use of ordinary empirical descriptive vocabulary.

---

76 Many everyday uses of modal vocabulary to qualify claims suppress the premises from which the claim implicitly is taken to follow, and so court the danger of countenancing the modal fallacy that would infer from p and □(p→q) to □q. Thereon hangs a tale.

77 CDCM § 108.

II. A Modal Realism

6. This sketch of a program for extending the Kant-Sellars tradition of modal expressivism raises a myriad of questions, some of detail, others more substantial. Rather than beginning to fill in that sketch by addressing some of those questions, I want to confront the ideas that motivate it with a different set of intuitions: those that motivate a robust modal realism. By "modal realism" I mean the conjunction of the claims that:

MR1) Some modally qualified claims are *true*.

MR2) Those that are state *facts*.

MR3) Some of those facts are *objective*, in the sense that they are independent of the activities of concept-users: they would be facts even if there never were or never had been concept-users.[78]

There are strong reasons to endorse all three of these claims. As to the first, physics tells us things such as: "Two bodies acted upon only by gravitational forces necessarily attract one another in direct proportion to the product of their masses and in inverse proportion to the square of the distance between their centers of mass." I take it this claim, for instance, is true. Even if it is not, I take it that *some* claims of this form, purporting to state laws of nature, do, in fact, state laws of nature. Denying this brings one into direct contradiction with the empirical sciences themselves. Supporting such a position would require a strong argument indeed. For the empirical sciences are in the business of making subjunctive and counterfactual-supporting claims. That is, they offer not only *descriptions*, but *explanations*. Indeed, the descriptions they offer are essentially, and not just accidentally, available to figure in explanations of other descriptions.

78 Of course, this is itself a modal claim, expressed counterfactually in the subjunctive mood. That fact is not problematic in the current context. One upshot of the previous discussion is that *any* description of how things objectively are implicitly involves modal commitments.

The second claim is, I think, true in virtue of the definition of 'fact'. A fact, Frege says, is a thought that is true.[79] He means 'thought' in the sense of something thinkable, not in the sense of a thinking, of course. For there can be unthought facts. On this usage, it is alright to say that facts make thoughts or claims true only in the sense that facts make acts of thinking and claiming true. For the facts just *are* the true thinkables and claimables. Wittgenstein is appealing to this way of using 'fact' when he says: "When we say, and mean, that such-and-such is the case, we—and our meaning—do not stop anywhere short of the fact; but we mean: this—is—so."[80] On this usage, if there are true modal claims—in the sense of true modal claimables, or modal claimings that are true in that they are claimings of true claimables—then there are modal facts. Modal facts are just facts statable using modal vocabulary, as physical facts are facts statable using physical vocabulary, nautical facts are facts statable using nautical vocabulary, and so on.

The third claim is perhaps the most controversial of these three platitudes. But I think the same principle I implicitly invoked in talking about the first claim underwrites it. Physics tells us that the current laws of nature were already laws of nature before there were human concept-users. And, although it does not specifically address the issue, it is clearly committed to the claim that the laws would have been the same even if there never had been concept-users. Indeed, many of the laws of nature (including all the Newtonian ones) exhibit a temporal symmetry: they hold indifferently at all times. So they are independent of the advent, at some particular time, of concept-users. And one of the mainstays of physics over the last century—substantially contributing to its distinctive conceptual shape—is the result of the Noether theorem that tells us (entails) that that this fundamental temporal symmetry is mathematically equivalent to the physical principle of conservation of energy.[81] Denying MR3 is denying the temporal symmetry of laws of nature. And the theorem tells us that that means denying the conservation of energy. While there are reasons from the bleeding edge of physics to worry about the universal truth of the principle of conservation of energy, those considerations are

79 In "The Thought" [ref.].

80 *Philosophical Investigations* [ref.] §95.

81 Cf. for instance Nina Byers (1998) "E. Noether's Discovery of the Deep Connection Between Symmetries and Conservation Laws." in *Proceedings of a Symposium on the Heritage of Emmy Noether*, held on 2–4 December, 1996, at the Bar-Ilan University, Israel.

irrelevant in the current context: they do not stem from the presence or absence of concept-users in our world. I conclude that one cannot deny MR3 without taking issue with substantial, indeed fundamental, empirical issues in physics.[82]

There is another line of argument to the conclusion that commitment to modal realism is implicit in commitment to a corresponding realism about claims expressed using ordinary empirical descriptive vocabulary. It will make clearer the relation between one kind of alethic modality and conceptual content. We can begin with a platitude: there is some way the world objectively is. How it objectively is must be discovered by empirical inquiry, and sets a semantic and epistemic standard for assessment of the correctness of our descriptive claimings as potential expressions of knowledge. The question is how to understand the relation of modal facts (if any) to how the world objectively is as describable (at least sometimes) in non-modal empirical descriptive vocabulary. One might ask a supervenience question here, but the line of thought I am concerned with goes a different way. It asks what modal commitments are implicit

82 I offer a different argument for this same conclusion (not specifically for the modal case, but for a more generic one that comprises it) in Section V of Chapter Five of *Perspectives on Pragmatism*: *Classical, Recent, and Contemporary*, forthcoming from Harvard University Press:

There were no true claimings before there were vocabularies, because there were no claimings at all. But it does not follow that there were no true claimables. In fact, we can show that we ought not to say that. Here is an argument that turns on the grammatical transformations that "It is true that…" takes.

Physics tells us that there were photons before there were humans (I read a lot about them in Stephen Weinberg's account of the early history of the universe, *The First Three Minutes* [New York: Basic Books, 1988], for instance). So if before time V there were no humans, so no vocabularies, we do not want to deny that

1. There were (at time pre-V) photons.

We can move the tense operator out front, and paraphrase this as:

2. It was the case (at time pre-V) that [there are photons].

By the basic redundancy property of 'true', we can preface this with "It is true that…":

3. It is true that [It was the case (at time pre-V) that [there are photons]].

Now we can move the tense operator out to modify the verb in "It is true that…":

4. Was[It is true (at time pre-V) that [there are photons]]

This is the key move. It is justified by the observation that *all* sentential operators can be treated this way, as a result of deep features of the redundancy of 'true'. Thus one can transform "It is true that Not[*p*]," into Not[It is true that *p*], "It is true that Possibly[*p*]," into "Possibly[It is true that *p*]," and "It is true that Will-be[*p*]," into "Will-be[It is true that *p*]." But now, given how the tense operators work, it is straightforward to derive:

5. It was true (at time pre-V) that [there are photons].

And again invoking the features that make 'true' redundant, we get:

6. It was the case (at time pre-V) that [It is true that [there are photons]].

These uniformities involving the interaction of 'true' with other sentential operators tell us we are committed by our use of those expressions to either deny that there were photons before there were people—which is to deny well-entrenched deliverances of physics—or to admit that there were truths about photons before there were people to formulate them.

already in the idea of an empirically describable world. It focuses on the *determinateness* of the way things objectively are.

To talk about how things objectively are as determinate is to invoke a contrast with how they are not. This idea is summed up in the Spinozist (and scholastic) principle *omnis determinatio est negatio*. This thought is incorporated in the twentieth-century concept of information (due to Shannon[83]), which understands it in terms of the partition each bit establishes between how things are (according to the information) and how they are not. But there are different ways we might follow out this idea, depending on how we think about the sort of negation involved. What I'll call the "Hegelian" model of determinateness insists that it must be understood as what he calls "exclusive" [ausschließend] difference, and not mere or "indifferent" [gleichgültig] difference.[84] **Square** and **circular** are exclusively different properties, since possession by a plane figure of the one excludes, rules out, or is materially incompatible with possession of the other. **Square** and **green** are merely or indifferently different, in that though they are distinct properties, possession of the one does not preclude possession of the other. An essential part of the determinate content of a property—what makes it the property it is, and not some other one—is the relations of material (non-logical) incompatibility it stands in to other determinate properties (for instance, shapes to other shapes, and colors to other colors). In fact, Hegel's view is that determinateness is a matter of standing in relations of material incompatibility (his "determinate negation") and material consequence (his "mediation") to other determinates. We might think of these as related by the principle that one property, say **metallic** is a consequence of another, **copper**, in case everything incompatible with being metallic (say, being a mammal) is incompatible with being copper. A property possession of which rules out possession of *no* other properties, and has as a consequence possession of no others, is in so far such *in*determinate.

One reason to endorse this Hegelian conception of determinateness is that it is required to underwrite what might be taken to be an essential aspect of the structural difference between the fundamental ontological categories of *object* and *property*. Aristotle had already pointed out a

83 Claude E. Shannon and Warren Weaver: *The Mathematical Theory of Communication.* The University of Illinois Press, Urbana, Illinois, 1949. ISBN 0-252-72548-4

84 The rubric 'Hegelian' here is tendentious, and liable to be alarming. More seriously, it is liable to be unhelpful. For now, treat it as a mere label. I will say what I mean by it—give it some content—as we go along.

fundamental asymmetry between these categories. It makes sense to think of each property as coming with a *converse*, in the sense of a property that is exhibited by all and only the objects that do *not* exhibit the index property. **Has a mass greater than 5 grams** is a property that has a converse in this sense. But it does *not* make sense to think of *objects* as coming with converses, in the analogous sense of an object that exhibits all and only the properties that are *not* exhibited by the index object. This is precisely because some of those properties will be incompatible with one another. Thus my left leg has the properties of not being identical to Bach's second Brandenberg concerto and not being identical to Gottlob Frege. Its converse, if it had one, would have to have the properties of being identical to both.

It should be clear that to take the objective world to be determinate in the Hegelian sense—so, to consist of objects and their properties and relations in the Aristotelian sense, and for those properties and relations to exhibit the structure of determinable families of determinates—is to be committed to modal realism. For Hegelian determinateness requires that there be facts about what properties and states of affairs are materially incompatible with which others, and about what material consequential relations they stand in to which others. The determinateness of the fact that this coin is copper consists in part in its being incompatible with the coin being silver and its having as a consequence that it conducts electricity—that is, with its being *necessary* that it is not silver, *possible* that it is green, and *necessary* that it conducts electricity.[85] Metallurgists discover these modal facts as part of the same kind of empirical inquiry through which they discover that this coin is in fact copper. A world without modal facts would be an indeterminate world: a world without objects in the Aristotelian sense, and without properties in the sense that admits a determinate-determinable structure.

In laying out Sellars's views I registered that he thinks of what he called the "causal modalities" as characterizing the inferential relations that articulate the contents of empirical descriptive concepts. If we go back to what Hegel made of Kant's views of modality and conceptual content, we find a notion of conceptual content that can help us better understand how this kind of modality can be understood as a *conceptual* modality. On this conception, to be *conceptually* contentful just is to stand in modally robust relations of material consequence and

85 Of course there are various provisos that would have to be added to make these claims strictly true, since copper can be alloyed with silver, and so on. I ignore these complications, as beside the point I am after.

incompatibility (what Hegel calls relations of "mediation" and "determinate negation"). This is a resolutely non-psychological sense of 'conceptual'. For it makes no reference to concept-*use*—to the *application* of concepts by anyone at all. So if there are laws of nature according to which some properties are incompatible with others (*can*not be exemplified by the same object at the same time) or have others as their consequences (if one is exhibited by an object, the other *must* be) then the world as it is objectively, independently of the activity of any knowing and acting subjects, is conceptually articulated. Empirical inquiry is at once the job of determining what judgments are true and what concepts are correct—that is, what really follows from what and what really precludes what. Linguistic terms can *express* concepts, by being used so as to undertake commitments as to what follows from what and what precludes what. But the concepts they express are in no sense *products* of that concept-applying activity.

As we saw, Sellars insists that it is standing in such relations that makes empirical descriptive vocabulary genuinely *descriptive*, that is, expressive of descriptive *concepts*, rather than merely functioning as reliably differentially responsively elicited *labels*. And we have seen that the sort of modal realism I have been sketching has as one of its consequences that empirical descriptive properties and states of affairs stand to one another in relations of material consequence and incompatibility. So Hegel offers us definitions of what it is to be *determinate* and to be *conceptually articulated*, according to which to take the objective world to be determinate is to take it to be *modally* articulated and to be *conceptually* articulated. That is, it commits one both to modal realism and to conceptual realism: the view that the objective world is modally, and *so* conceptually structured, quite apart from its relations to us.

III. Together

7. The core of the modal realism I have just sketched consists of some claims that express philosophical common sense: there are laws of nature, events sometimes causally necessitate others, there is a determinate way the world objectively is, and its being that way rules out (excludes the possibility) of its being some other ways. These are commitments to which any philosopher ought to want to be entitled. They should be contested only under theoretical duress by exceptionally weighty and compelling arguments.

But what is the relation between this kind of modal *realism* and the modal *expressivism* I talked about in the first part of this talk? There the expressive role characteristic of modal vocabulary was identified as making explicit the material inferential and incompatibility relations in virtue of which ordinary empirical descriptive (OED) vocabulary expresses the content that it does. This expressive role was distinguished from that of the ground-level empirical descriptive vocabulary, whose principal job it is to say how things objectively are. There is no further vocabulary to which OED vocabulary stands in the same semantically explicative relation as alethic modal vocabulary stands to it.[86] The core of this version of modal expressivism lies precisely in the distinction it insists on between the expressive role distinctive of modal vocabulary and that of vocabulary whose job is describing the world, at least in the narrow, paradigmatic sense in which OED vocabulary describes the world. Modal realism says that modal vocabulary *does* describe the world, does say how things are. So are these two lines of thought simply incompatible? Are we obliged to choose between them?

I think that the modal expressivism of Part I and the modal realism of Part II are not only compatible, but that that account of the *expressive* role distinctive of modal vocabulary is just what is needed to understand the central claims of modal *realism*. The expressivism complements, rather than conflicting with, the realism about the use of modal concepts. How is such a reconciliation to be understood? The first step is to see that modal expressivism (ME)

86 This is the expressive role of being *elaborated from* and *explicative of* the use of OED vocabulary. It is what in *Between Saying and Doing* I call "being LX" for that vocabulary.

makes claims about what one is *doing* in using modal concepts, while modal realism (MR) makes claims about what one is *saying* by using modal concepts. ME says that what one is doing when one makes a modal claim is endorsing an inference relating descriptive concepts as subjunctively (including counterfactually) robust, or treating two descriptive concepts as incompatible. MR says that when one does that, one is saying (claiming) *that* possession or exhibition of one empirical property is a consequence of, or is incompatible with, possession or exhibition of another. The claim that ME and MR are *compatible* is the claim that one can *both* be *doing* what ME says one is doing in applying modal vocabulary *and* be *saying* what MR says one is saying by doing that. The claim that they are *complementary* is the claim that an important way to understand what one is *saying* by making modal claims is precisely to think about what one is *doing* by making them.

According to this way of understanding the relations between ME and MR, the claims of modal expressivism are made in a *pragmatic* metavocabulary for modal vocabulary: that is, a vocabulary suitable for specifying the practices, abilities, and performances that make up the *use* of modal vocabulary. And the claims of modal realism are made in a *semantic* metavocabulary for modal vocabulary: that is, a vocabulary suitable for specifying the *meanings* or conceptual *contents* expressed by modal vocabulary. What we have here is an instance of the general question of how to understand the relations between these two complementary aspects of concept application in claims: the use of the concepts and their meaning or content, what one is doing by applying them and what one is saying by applying them.

Modal expressivism says that what one is doing in making modal claims is not the same thing one is doing in making claims using ordinary empirical descriptive vocabulary. For in the former case, but not the latter, one is (perhaps *inter alia*) committing oneself to subjunctively robust inferential-and-incompatibility relations among descriptive concepts one is not in general thereby applying. Modal realism says that in making modal claims one is saying how things objectively are, describing the objective world. Reconciling these claims requires specifying a sense of "describing" or "empirical fact-stating" that is broader than that applicable to the primary use of OED vocabulary, but still sufficiently akin to it that the broader sense applicable to modal claims and the narrower sense applicable show up as species of a recognizably descriptive genus.

8. Here is a suggestion: A broader sense of "fact-stating" and "description" that is not yet so promiscuous as the declarativist candidate that treats all declarative sentences as descriptions is defined by the dual requirements of *semantic government* of claimings by facts and *epistemic tracking* of facts by claimings.

By "semantic government" I mean that descriptive claims are subject to a distinctive kind of ought-to-*be* (related only in complicated ways to the ought-to-*do*s that Sellars contrasted them with). It ought to be the case that the content of a descriptive claiming stands in a special relation, which we might as well call "correspondence," to a modal fact, which it accordingly purports to state (and in case there is such a fact, succeeds in stating). In virtue of that semantic norm, claimings are answerable for their correctness (accord with that norm) to facts. The underlying thought here is that what one is talking *about* is what exercises a certain kind of *authority* over what one says; what one says is *responsible to* what one is talking about, in a way that is characteristic of this relation as *semantic*. What one is talking about provides a standard for the assessment of what one says.

What is the nature of the correspondence that the norm enjoins? The contents of possible claimings are articulated by relations of material consequence and incompatibility to the contents of other potential claimings. These notions are themselves specifiable in a *deontic normative* pragmatic metavocabulary: committing (or entitling) oneself to one claim can commit (or entitle) one to others, and can preclude entitlement to still others. The contents of facts and possible facts are also articulated by relations of material consequence and incompatibility to the contents of other possible facts. In this case, these notions are specifiable in an *alethic modal* semantic metavocabulary: the obtaining of one fact has the obtaining of others as a necessary (that is, subjunctively, including counterfactually, robust) consequence, makes others possible, and rules out still others as not possible. Normative semantic government of claimings by facts says that it ought to be the case that there is a fact whose content is articulated by objective modal relations of material consequence and incompatibility that line up with the subjective (in the sense of pertaining to knowing and acting discursive subjects) normative relations of material

consequence and incompatibility that articulate the content of a claiming. If that norm is not satisfied, the claiming does not live up to the standard provided by the fact it purports to state.[87]

Where semantic government of claiming by facts is a (deontic) *normative* matter, epistemic tracking of facts by claimings is a(n) (alethic) *modal* one. It is a matter of the subjunctive and counterfactual robustness of the conceptual content correspondence between facts and claims. The tracking condition holds just insofar as the subjunctive conditional "If the fact were (or had been) different, the claiming would be (or would have been) correspondingly different," is true. Insofar as this condition holds, there is a *reliable* correspondence between the contents of facts and the contents of claimings. That is to say that the inference from a claim about the content of a claiming to the content of the corresponding fact is in general a good one.

9. I think it is a fundamental mistake to try to do all the work of done by the concept of semantic government with that of epistemic tracking, as for instance Fodor and Dretske do. What goes missing is the fine structure of the crucial interaction between activities on the part of the claiming subject, expressed in a deontic normative pragmatic metavocabulary, and how it is with the objects and facts those claims are about, expressed in an alethic modal semantic metavocabulary, and how the two sides stand in both normative relations of semantic government and modal relations of epistemic tracking. It is precisely in these intricate relations that the complementary character of modal expressivism and modal realism becomes visible.

When the two requirements of semantic government and epistemic tracking are satisfied, it makes good sense to think of the claimings in question as fact-stating and descriptive. They purport to say how things are with what they in the normative sense of semantic government *about*. The actual applications of the vocabulary in question, no less than their normative status as correct or not, are epistemically *responsive* to and *controlled* by the corresponding facts.

This is also evidently true also of modal vocabulary, supposing we grant the dual claims of modal realism and modal expressivism. For modal *expressivism* tells us that modal vocabulary makes explicit normatively significant relations of subjunctively robust material

87 The concept of propositional content as what is articulated by relations of material consequence and incompatibility is a development of the Fregean metaconceptual semantic dimension of *Sinn*, while the normative relation of aboutness between objective facts and subjective commitments is a development of his metaconceptual semantic dimension of Bedeutung.

consequence and incompatibility among claimable (hence propositional) contents in virtue of which ordinary empirical descriptive vocabulary *describes* and does not merely *label*, *discriminate*, or *classify*. And modal *realism* tells us that there are modal facts, concerning the subjunctively robust relations of material consequence and incompatibility in virtue of which ordinary empirical descriptive properties and facts are determinate. Together, these two claims give a definite sense to the possibility of the correspondence of modal claimings with modal facts. If we can then say what it is for a norm of semantic governance to be instituted and the modal fact of epistemic tracking to be achieved, the descriptive, the fact-stating character of modal vocabulary according to ME and MR will have been made intelligible.

It is a consequence of the version of Kant-Sellars modal expressivism that I outlined in Part I that instituting normative semantic government of modal claims by modal facts, and of achieving modal epistemic tracking of modal facts by modal claims must be an aspect of the process of instituting semantic government of ordinary empirical descriptive claims by the facts they state, and of achieving epistemic tracking of those facts by ordinary empirical descriptive claims. For the essence of that view is that what is expressed *explicitly* (that is, put in claimable, propositional form) by the use of modal vocabulary is already *implicit* in the norms governing the use of OED vocabulary. Determining and applying descriptive concepts inevitably involves committing oneself as to the subjunctively robust inferential and incompatibility relations they stand in to one another. Rectifying concepts, determining facts, and establishing laws are all projects that must be pursued together. Empirical evidence bears on all of the semantic, epistemic, and explanatory tasks at once, or it bears on none of them.

If that is right, then modal claims (and the concepts that articulate them) exhibit semantic government by and epistemic tracking of facts no less than ordinary empirical descriptive ones do. Far from being incompatible with this fundamental modally realistic claim, modal expressivism is just what is needed to make it intelligible. By showing how the use of modal concepts and the use of ordinary empirical descriptive concepts are inextricably bound up with one another, modal expressivism also shows itself and modal realism as two sides of one coin.

IV. Again

10. I have argued that modal realism and the right kind of modal expressivism belong together. The tendency to understand views of this kind as incompatible alternatives—to take the sense in which modal vocabulary plays, as Sellars put it a "metalinguistic" expressive role relative to ordinary empirical descriptive vocabulary to rule out the possibility of its being also fact-stating and descriptive of something other than language use—is the result of failing to attend to the distinction beween *pragmatic* and *semantic* metavocabularies.

11. I have finished my argument. But I want to close with a lagniappe, indicated in the final word of my title. Why claim, as that title does, that the result of this story is to put modal expressivism and modal realism together *again?* Why should the story be thought of as recounting a *re*union? The answer I want to leave you with is this: It is because we've seen something very like this constellation of metaconceptual commitments before. I started my story with Kant, and that is where I want to end it. Claiming that one should be a *pragmatic* modal expressivist (an expressivist about what one is *doing* in applying modal vocabulary) but a *semantic* modal realist (a realist about what one is *saying* in applying modal vocabulary) is, I think, recognizably a development and a descendant, for this special but central case, of Kant's claim that one should be a *transcendental idealist,* but an *empirical realist.* That is what I mean by saying that the view I have been presenting puts modal expressivism and modal realism together *again*.

END of Lecture 5