Introduction
Metavocabularies of Reason

Robert B. Brandom

0.1 Two Traditions in the Philosophy of Language and Two Kinds of Metalanguage

One of the principal distinguishing features of philosophical thought over the past one hundred and fifty years is its pursuit of the strategy of addressing philosophy's perennial concern with the nature of understanding or reason in general by investigating *language*, thought of as its essential, characteristic embodiment and expression. That guiding idea has been developed in two quite different directions, giving rise to two nearly independent and often antagonistic currents of thought. The first, dominant, better worked-out tradition focused on logic, and later, also formal semantics, as perspicuous mathematical metalanguages in which to articulate conceptual contents and their rational relations to one another. The other tradition focused rather on language as a kind of social practice, thought of as a prominent feature of the natural history of a certain kind of being: language as *parole* rather than *langue*, in de Saussure's terms.

The logico-semantic formalist tradition can be roughly delineated by tracing a central strand that leads from Gottlob Frege through Bertrand Russell and the Wittgenstein of the *Tractatus Logico-Philosophicus*, to Rudolf Carnap, Alfred Tarski, Saul Kripke, David Lewis, and Kit Fine. Its paradigm is the language of mathematics, from which it takes the model of justifying reasons as apodeictic proofs. The founding achievement of this tradition is Frege's invention of a logical calculus, itself a regimented mathematical vocabulary, adequate to express the reason relations and conceptual contents of arithmetic. That feat inspired the project of using the powerful expressive tools and explicative methods of the new logic to codify reasons and conceptual contents generally, with hitherto unapproachable precision and perspicuity. Beginning with Tarski's model theory (further developed, and packaged for and marketed to philosophers by Carnap and Quine), formal semantic metalanguages joined formal logical ones in the mathematical metaconceptual toolkit defining this

tradition. The expressive power of Tarskian model theory was substantially enhanced by its development in the form of possible worlds semantics, and then again in its further generalization into contemporary truth-maker semantics. In each case, the first application and proof of concept of the new semantic metalanguages was their demonstrable adequacy in providing a semantics for specifically *logical* vocabularies (in the case of possible worlds semantics, for modal logics). It was then shown how to extend those achievements to provide, first extensional (Tarski, Quine), then intensional (Lewis), and hyperintensional (Fine) semantics for *non*logical (and nonmathematical) expressions.

Avatars of the anthropological, natural-historical, broadly pragmatist tradition include the classical triumvirate of American pragmatists, Charles Peirce, William James, and John Dewey, and also the Heidegger of *Being and Time*, the Wittgenstein of the *Philosophical Investigations*, Richard Rorty, and such contemporary figures as Charles Travis and Huw Price. Where the formalist tradition is oriented by a conception of understanding and reasons as codified in artificial logical calculi and semantic metalanguages, the pragmatist tradition looks instead directly to natural languages, thought of as social practices and forms of life. In place of the exclusively monological character of reasoning as deriving, modeled on proof, characteristic of the other tradition, understanding shows up in this tradition as a social achievement, and reasoning as essentially dialogical: a matter of discursive practices of giving and asking for reasons, defending and challenging claims that amount to taking up positions in a contestable, public, normative space.

These two general philosophical approaches to language have been pursued to a remarkable extent independently of one another, rather than in conversation with each other. That fact both engenders and is supported by the impression that they are sufficiently antithetical that philosophers interested in language are obliged to choose between them, and to align themselves with one or the other camp. That unhappy impression is gainsaid by influential figures such as Davidson, Putnam, Sellars, Dummett, and Stalnaker, who, like Quine (and even Lewis), are distinguished by their labors in both traditions, and (unlike the time-slices of Wittgenstein) in developing and promoting their confluence. The two traditions ought by rights to be understood as focusing on different aspects of language: roughly, on the *meanings* of linguistic expressions, and on their *use*. In suitably broad senses, we might understand semantics as the study of meaning, and pragmatics as the study of use or discursive practices and abilities. So understood, semantics (even a semantics inspired by and paradigmatically applicable to logic) and pragmatics show up as complementary theoretical endeavors. The goal should be to synthesize semantic and pragmatic theories, so as better to understand the relations

between meaning and use, as two essential aspects of a single linguistic phenomenon.

That aspiration is compatible with acknowledging that there might still be good arguments for adopting a semantics-first or a pragmatics-first order of explanation. The thought that the norms that function as standards for assessments of the correct use of expressions must be sensitive to the meanings (for instance, the truth conditions) of the expressions used speaks for a semantics-first order of explanation. And the thought that all there is to confer meanings or contents on linguistic expressions is the (actual or proper) use of those expressions speaks for a pragmatics-first order of explanation. Perhaps the combination of those thoughts recommends rather a more balanced view that eschews claims of explanatory priority in favor of understanding each aspect as in principle intelligible only in terms of its relation to the other.

We pursue here a synthetic metalinguistic strategy. We think of semantics and pragmatics as theoretical undertakings pursued by talking about language in two different kinds of metalanguage: semantic metalanguages and pragmatic metalanguages (or, as we will say: metavocabularies). They make it possible to express, to make explicit, different aspects of a single phenomenon. Further, we think that powerful new insights into the nature of language can be provided by understanding that phenomenon as essentially, and not just accidentally, having aspects that can be made explicit by just the right semantic and pragmatic (and, indeed, logical) metavocabularies.

The lesson that emerges, we will argue, is a kind of discursive or linguistic *rationalism*. Language becomes visible as at base the medium of reasons, and reasoning as the beating heart of language. On the side of pragmatics, the fundamental speech act is that of making claims. The basic speech act of making claims, asserting, is to be understood in terms of practices of defending and challenging those claims, by making other claims that have the practical significance of giving reasons for and against them. Understanding claiming this way provides a path to understanding the claimable contents expressed by declarative sentences in terms of the role they play in relations of being a reason for or against— what we will call "reason relations." On the side of semantics, worldly represented states show up as what determine the reason relations of consequence and incompatibility that the sentences whose truth-makers and falsifiers they are stand in to one another: their roles in reason relations. By understanding the common topic that semantic and pragmatic metalanguages articulate aspects of, not just under the vague rubric of "language," but more specifically as the implicit reason relations that distinguish *discursive* practices as such, we can better understand not only the relations between the meaning and the use of linguistic expressions,

but also the relations between truth (the central concept of traditional semantics) and justification (the central concept of pragmatics, according to linguistic rationalism), in the form of practices of defending claims by giving reasons for them and challenging claims by giving reasons against them. One manifestation of that opposition between truth-based and justification-based approaches is the distinction between model-theoretic and proof-theoretic formal metavocabularies for codifying reason relations. We will have a lot to say about the relations between them in the rest of this book.

The multifaceted metalinguistic rationalism we present in the chapters that follow offers a new way of understanding why and how what is expressed by declarative sentences is both what can be semantically evaluated as true or false and what can be pragmatically accepted or rejected. It articulates a conception of sentential *conceptual content* in terms of the functional roles sentences play in virtue of standing to one another in reason relations of consequence and incompatibility, rational inclusion and exclusion, which show up in different, though intimately related guises, from the perspectives afforded not only by semantic and pragmatic metalanguages, but also by logical and metalogical ones. There is, of course, much more to meaning and to use than what is precipitated in conceptual content as we consider it. But seeing how the aspects of language that are addressed by logical, semantic, and pragmatic metalanguages can be understood as providing perspectives on the common topic of reason relations and roles with respect to them opens up avenues for reconciling and synthesizing the often apparently divergent things these vocabularies make it possible for us to say and think, about what we are saying and doing when we are talking and thinking.

0.2 Reason Relations, Vocabularies, and Metavocabularies

At the core of this book, then, is the rationalist explanatory strategy of understanding the nature of language in terms of what we call "reason relations." As addressed here, that is a genus with two principal species: implication and incompatibility. They correspond to being a reason for and being a reason against. One of our overarching aims is to offer a detailed, multifaceted account of reason relations as such. A subsidiary allied task is more precisely to delineate the object of study, by showing how reason relations appear from the quite different perspectives afforded by the vantage points of logic, semantics, pragmatics, and a new formal apparatus for articulating the conceptual roles declarative sentences play in virtue of standing to one another in relations of rational inclusion and exclusion. The common topic that shows up in various guises when we triangulate (really, quadrangulate) on it by considering what can be said about it in

these four disparate technical kinds of vocabulary is in many ways both familiar and recognizable, but also subtly yet importantly different from what is typically discussed under the heading of "reasons."

A closely related term of art is "vocabulary." We use it in a technical sense, to mean a lexicon or set of declarative sentences, together with an implication relation and an incompatibility relation defined on those sentences. To begin with, we can think of an implication relation as holding between a set of sentences that are its premises and a single sentence that is a conclusion that follows from, is a consequence of, or is implied by those premises. An incompatibility relation holds between a set of premises and a further sentence that those premises exclude, or rule out, or are incompatible with. The elements of the lexicon are repeatable sign designs of any sort. By calling them (declarative) "sentences" we just mean that they are what in the first instance stand to one another in reason relations of implication and incompatibility. (The reason relations among predicates—not considered in our treatment, which remains resolutely at the level of sentences—are derivative from these.) In virtue of standing to one another in reason relations of implication and incompatibility, what thereby count as declarative sentences express conceptual contents—specifically *propositional* contents. Those contents can be thought of as the functional roles the sentences play in constellations of implications and incompatibilities. Sentences express what can both serve as and stand in need of reasons (for and against, positive and negative).

This characterization is very abstract. Whatever plausibility there is to characterizing the elements of the lexicon of a vocabulary as "declarative sentences," and the roles they play as "conceptual or propositional contents" depends at least on the relations defining what we are calling a "vocabulary" really being reason relations of implication or consequence and incompatibility. (Even then, the connections being asserted between sentencehood and propositional contentfulness, on the one hand, and reason relations, on the other are substantive, potentially controversial, and so in need of specification, motivation, and justification. *Solvitur ambulando*.) According to this order of explanation, the key question is: what do we mean by talk of reason relations of implication and incompatibility? In virtue of what does something deserve to count as a consequence or incompatibility relation? On the approach we are pursuing, an answer to this question will ramify into answers to corresponding questions about sentences and conceptual content.

The idea is to identify reason relations in terms of the various vocabularies that can be used to specify them. Because these are vocabularies for talking about (the reason relations of) other vocabularies, they are *meta*vocabularies. Because it is in particular the reason relations of base vocabularies that they address, we can call them

rational metavocabularies—by contrast to metavocabularies that express other features of the lexicon: syntactic, phonological, or etymological metavocabularies, for instance. Two different strategies for explaining what reason relations are—corresponding to two different philosophical traditions of thought about language—are pursued in two different kinds of rational metavocabulary: semantic and pragmatic.

Semantic metavocabularies explain what is expressed by the reason relations of empirical base vocabularies by saying what features of the world those vocabularies capture in their consequence and incompatibility relations. The sentence "The coin is made of pure copper" implies the sentence "The coin would melt at 1085°C." because and in the sense that pure copper *necessarily* melts at 1085°C., and it is incompatible with the sentence "The coin is an electrical insulator," because it is *impossible* for pure copper not to conduct electricity. Semantic metavocabularies explain reason relations of implication and incompatibility by specifying what the sentences that stand in those relations *mean*, in the sense of how the world must be for what they say to be true. The sentences stand to one another in relations of implication and incompatibility because the objective states of affairs that are their semantic truth conditions stand to one another in modally robust relations of necessitation and noncompossibility. For sentences to be part of a vocabulary, and not just a lexicon, is for them to express meanings in the form of worldly truth conditions that are themselves related to one another by consequence and incompatibility relations that are specifiable in an alethic modal semantic metavocabulary.

Pragmatic metavocabularies explain what is expressed by reason relations of base vocabularies by saying what features of the discursive practices of *using* those sentences it is, in virtue of which practitioners count as practically taking or treating the sentences *as* standing to one another in relations of implication and incompatibility. Pragmatic metavocabularies make it possible to *say* what it is that language users *do* in virtue of which they are properly to be understood as practically taking or treating some sentences as implying others in the sense of taking assertion or acceptance of the premises as providing reasons for asserting or accepting the conclusions, and practically taking or treating some sentences as incompatible with others in the sense of taking assertion or acceptance of the premises as providing reasons against asserting or accepting the conclusions. Reason relations show up from the expressive perspective provided by pragmatic metavocabularies as normative standards for assessment of the correctness of rational defenses of and challenges to claims, made by offering other claims as reasons for or reasons against those claims. Using lexical items *that* way—as specified in the pragmatic metavocabulary—is using them *as* the sentences of a vocabulary.

These two kinds of metavocabulary are evidently different in many important ways. Semantic metavocabularies talk about the *meanings* expressed by the sentences of a base vocabulary, in the sense of how those sentences represent the objective world as being. Pragmatic metavocabularies talk about the *use* of sentences in virtue of which they express those meanings. As we will see later in much more detail, to do their job properly, semantic metavocabularies must use alethic modal vocabulary to make claims about what states and combinations of states of the world the base vocabulary talks about are and are not possible. To do their job properly, pragmatic metavocabularies must use deontic normative vocabulary to make claims about what acts, practical attitudes, and combinations of them are and are not appropriate, and what other acts and attitudes would and would not entitle an interlocutor to them. What can be said in alethic modal terms is substantially and importantly different from what can be said in deontic normative terms. The one concerns features of the objective world, the other features of the practice of discursive subjects. These are the two poles of the intentional nexus that links knowers and the known, minds and the world they understand and act in, representings and what is represented. We want to understand both kinds of thing, and the important relations between them. That is a philosophical challenge because of the substantial differences between them.

In claiming that both pragmatic and semantic metavocabularies can be regarded as *rational* metavocabularies, though, we are claiming that in spite of these differences of subject-matter, concern, and metaconceptual resources, they do have at least one common topic. Among the things they both discuss and seek to explain in their own terms are reason relations. This claim of an overlap in topics is a strong, and perhaps surprising one. Consider for the moment the example of incompatibility. The claim is that when the meaning-theorist understands the incompatibility of p and q as a matter of the fusion of any objective state that is a truth-maker of p with any truth-maker of q being an impossible state in an alethic modal sense, and the use-theorist understands the incompatibility of p and q as a matter of the position an interlocutor is in when accepting both p and q as being normatively out of bounds—one the interlocutor cannot be entitled to—while they are clearly using "incompatible" in different *senses*, there is a common *referent*, a "reason relation," that both are picking out and talking about, in their respective idioms. In the terms used above to introduce the idea of reason relations, we propose to understand the alethic modal semantic metavocabulary and the deontic normative pragmatic metavocabulary as offering different (meta)conceptual *perspectives* on a common object: the incompatibility of what is expressed by the declarative sentence p and what is expressed by

the declarative sentence *q*. Corresponding claims apply to reason relations of consequence or implication.

A sketch has been offered of how a kind of metalinguistic rationalism might serve to make mutually intelligible two traditions in the philosophy of language, to synthesize them by exhibiting them as articulating different dimensions of a unified phenomenon. The present point is that if the claim that it is possible to identify a rational structure common to what is expressed in pragmatic and semantic metavocabularies could be made out in detail, it would cast light on issues of much wider philosophical significance. For we can look at the relations between what is expressed in normative pragmatic and representational semantic metavocabularies in another way: as articulating the relations between the activities of talking and thinking, and what is being talked or thought *about*. This is the intentional nexus between subjects and objects, between mind and world, knowers and the known.

The American Pragmatists inherited from the German Idealists—who in turn inherited it from Romantic critics of the Enlightenment—the idea that the Cartesian tradition failed structurally, making itself a patsy for skepticism, by attempting to define subjects and objects independently of one another, and then later on facing the problem of how to bolt together things understood as having wholly disparate natures. (A distinction becomes a dualism when it is drawn in terms that make the relations between the distinguished items unintelligible.) The better strategy, they thought, was to start with a conception of intentionality as successful cognition (and action), and analyze the contribution of its subjective and objective aspects. One way to work out such a strategy begins with the thought that there is a kind of *structure* common to what normative pragmatic metavocabularies make it possible to say about the practices of discursive subjects using declarative sentences to manifest practical attitudes and undertake commitments, on the one hand, and what representational semantic metavocabularies make it possible to say about the modal relations among matter-of-factual states of the world those sentences come to represent by being so used, on the other.

The further thought that that common structure deserves to be understood as a specifically *conceptual* structure—in the sense of that term that indicates that the roles sentences that are used to make, defend, and challenge assertible and deniable claims, and what makes those claimables true or false, stand to one another in relations of consequence and incompatibility—begins to develop this strategy in the direction of a *conceptual realism*. It appeals to a sense of "conceptual" that does not restrict that term to the products of thought. For the worldly version of the relations that articulate the structure we are calling "conceptual" are relations of necessity and impossibility whose existence owes nothing to the

activities of discursive practitioners. They are objective relations, specified in the alethic modal vocabulary used to state laws of nature, and more generally, to specify subjunctively robust relations. The overall idea is that relations that *in some sense* correspond to those objective relations on the "subjective" side of the activities and practical attitudes of speakers and thinkers are reason relations of consequence and incompatibility specified in the deontic normative modal vocabulary of consequential commitments and preclusion of entitlements. Since the objective and subjective "aspects" of the claimed common conceptual structure are those specified in vocabularies using two different sorts of modality, alethic and deontic, respectively, this view can be denominated a specifically *bimodal* conceptual realism. In addition, the version of bimodal conceptual realism we will offer is *hylomorphic* in a recognizably Aristotelian sense. For it identifies a kind of *rational form* that is understood as common to thoughts and things, both to the assertibles/deniables commitment and entitlement to which is at issue in speech acts of claiming, and defending and challenging claims, and to the worldly states that make such claimables true or false. The common conceptual structure is a *rational* form because, as we will argue, the relations of consequence and incompatibility that show up in different guises in a whole constellation of intimately interrelated metavocabularies—not just pragmatic and semantic, but also logical and metalogical formal vocabularies dedicated to characterizing conceptual roles as such—are those that in the end underwrite practices of reaso*ning*, by determining what is a reason for and against what.

One aim we are pursuing here, then, is to give a clear and definite sense to the term "conceptual realism," by means of a kind of metalinguistic functionalism. The genus of which the conceptual realism we explore is a species has been variously characterized by different philosophers. We take the view we develop to be a way of understanding what Frege means when he says "A fact is a thought that is true."[1] It is also one way of understanding the Tractarian claim that the world is the totality of facts, and that facts must be understood as what can be pictured in thought. So understood, it is a commitment Wittgenstein never relinquished, asserting in the *Investigations*: "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" (Wittgenstein, *PI* §95). John McDowell (1996) explores the same sort of conceptual realist view in *Mind and World* under the slogan "The conceptual has no outer boundary."

These are deep waters. These pronouncements by great philosophers are mentioned to indicate that the stakes are high for the enterprise of explicating any form of conceptual realism. Here is a sketch of how we go about it. One of the key arguments we appeal to in filling in this neo-Aristotelian metalinguistic bimodal conceptual realism is a technical result

presented in Chapter Four. It articulates a clear and precise sense in which semantic and pragmatic metavocabularies of the kinds invoked here can be understood as offering different perspectives on a common topic. Greg Restall and David Ripley have worked out what they call a "bilateral" normative pragmatic understanding of the turnstile that marks implication relations in multisuccedent sequent calculi. They interpret the claim that a set of sentences Γ implies a set of sentences Δ—in the notation we will use here, written as the sequent Γ ⊢∼ Δ—as meaning that the "position" one is in if one asserts or accepts all of Γ and denies or rejects all of Δ is normatively "out of bounds." (The account is called "bilateral" in virtue of treating the premise side of the turnstile differently from the conclusion side: taking the one to concern assertions and the other denials.) In the deontic normative vocabulary we will use to render this view in Chapter One, that constellation of *commitments* (to accept and reject) is one that one cannot be *entitled* to. The idea is that what discursive practitioners must *do* in order thereby practically to be taking or treating the sets of sentences Γ and Δ *as* standing to one another in the basic reason relation of implication is to treat the corresponding position as normatively inappropriate or out of bounds.

The Restall-Ripley bilateral normative pragmatic metavocabulary turns out to be related in surprising ways to what we take to be the most sophisticated contemporary heir of the representational metavocabularies of Tarskian model theory and later intensional semantics in terms of possible worlds (Lewis, out of Kripke, out of Carnap), namely Kit Fine's truth-maker semantic framework. It incorporates a flexible metaphysical vocabulary for specifying the objective universe that is available to be represented: it consists of "states," and all their mereological fusions. The states are divided into possible and impossible states. Possible worlds are rendered as mereologically maximal possible states. But there are also much more limited possible states, perhaps such as the frog being on the log or Pittsburgh being in Pennsylvania. In addition, there are many different impossible states, perhaps there being immortal mammals or faster-than-light particles. The representational content of declarative sentences is then understood in terms of assignments to them of sets of states as truth-makers and falsifiers. Global structural conditions on modally partitioned state spaces (for instance requiring that all the mereological parts of possibles states be possible) interact with conditions on assignments of truth-makers and falsifiers (for instance forbidding the truth-makers and falsifiers of logically atomic sentences to be overlapping sets).

We show below that if one defines semantic consequence in just the right way, a powerful, fruitful, and detailed isomorphism can be constructed relating truth-maker modal semantic metavocabularies and bilateral normative pragmatic metavocabularies.[2] The required definition

is that a set of sentences Γ semantically entails a set of sentences Δ just in case every fusion of truth-makers of all of Γ with falsifiers of all of Δ is an impossible state. (This is substantially different from any definition of consequence currently abroad in the truth-maker literature.) Even at this rough level of description, the analogy this establishes with the bilateral pragmatic definition of consequence is clear. For the Restall-Ripley normative pragmatics treats Δ as a consequence of Γ just in case the position one would be in by asserting all of Γ and denying all of Δ is out of bounds. Assertion and denial line up with truth and falsity, combinations of commitments (to accept and reject) in a position line up with fusion of truth-making and falsifying states, and normative out-of-boundness (preclusion of entitlement to the commitments incurred by those assertions and denials) of a compound practical position lines up with the modal impossibility of such a fusion state.

When Spinoza looked back on the relations between algebraic equations and geometric shapes on which Descartes modeled mind-world relations, he saw that the key feature distinguishing that new, more abstract notion of representation from earlier atomistic resemblance-based conceptions is the existence of a global isomorphism between the algebraic and geometrical vocabularies. Spinoza's slogan for the holistic insight that animated the representational revolution was "The order and connection of ideas is the same as the order and connection of things" (Spinoza, *Ethics* II, Prop. vii). The isomorphism between normative pragmatic and alethic representational semantic metavocabularies turns out to make possible in our setting a precise, tractable, and productive specification of that shared rational "order and connection." We think this is a good way to rationally reconstruct some central aspects of Aristotelian (and Scholastic) intelligible forms. This isomorphism is the core of our version of bimodal (deontic/alethic) metalinguistic conceptual realism.

0.3 Intrinsic Metavocabularies for Reason Relations and Conceptual Contents: Logic

The construction gestured at so far foreshadows an argument for understanding reason relations of consequence and incompatibility as constituting a structure common to representational meaning and to practical use, to truth-making and to justificatory practices, to the objective world talked about and to the activities of talking about it, to what is represented and to the representing of it. That these same reason relations show up from the two otherwise disparate perspectives afforded by (the right kind of) semantic and pragmatic metavocabularies offers some reason to think of those relations as central to language or discourse as such. It provides a metalinguistic functionalist path to what we have

called "linguistic rationalism." Such an approach is unusual, and so perhaps surprising in how it discerns rational forms amphibious between these different dimensions. For the sort of language philosophers in the analytic tradition are accustomed to think of as codifying "rational forms" is formal *logical* languages, rather than semantic or pragmatic metalanguages. We agree that logical vocabulary is indeed uniquely privileged among languages for expressing reasons. But it is different from the rational metavocabularies considered here in several important ways. Our conception of the expressive role characteristic of logical vocabulary is also different from received views. One can begin to get a sense of it by comparing and contrasting it with the metavocabularies already under discussion.

One important criterion of adequacy for both semantic and pragmatic metavocabularies as we understand them is that they offer expressive resources sufficient to provide *explanations* of the reason relations of arbitrary base vocabularies. They are able to say, each in its own distinctive idiom, both what it means for some sentences to stand to others in relations of implication or incompatibility and why they do: what further facts explain them standing in just those relations. Our preferred version of semantics offers, in effect, truth-makers for the claims that $\Gamma \# A$ (Γ is incompatible with A) and $\Gamma \mathrel{|\!\!\sim} A$ (Γ implies A) in alethic modal terms of the impossibility of fusions of truth-making states of Γ with truth-making states of A, and truth-making states of Γ with falsifying states of A, respectively— that is, in terms of how the sentences of Γ and A represent the objective world to be. Our preferred version of pragmatics specifies how one must use sentences in order thereby to count as practically *taking* or *treating* them as standing in relations of implication or incompatibility. It does that in deontic normative terms of constellations of commitments to accept and reject the claimables they express being improper, inappropriate, or "out of bounds."

In order to make available substantive accounts of reason relations, these two very different kinds of rational metavocabularies bring to bear significant conceptual resources that are not already contained in the base vocabularies to which they apply: mereological and alethic modal metaphysical concepts and semantic interpretation functions from sentences to sets of truth-making and falsemaking states, and bilateral concepts of acceptance and rejection and the normative assessment of such practical attitudes, respectively. Because both of these kinds of metavocabulary appeal to conceptual resources beyond those intrinsic to the base vocabularies of which they are metavocabularies, and do so in service not just of characterizing the reason relations of those base vocabularies but of explaining them, the sorts of semantic and pragmatic

metavocabulary we consider can be denominated "extrinsic-explanatory" rational metavocabularies.

That is not the only kind of metavocabulary for specifying reason relations with which we are concerned. In addition to *extrinsic-explanatory* rational metavocabularies, there are also *intrinsic-explicative* ones. This latter kind of metavocabulary for reason relations restricts itself to the conceptual resources supplied by the base vocabularies whose reason relations it characterizes, and is used to make explicit those reason relations and the conceptual contents they articulate, rather than to explain why they are as they are, or what it is for them to be as they are. The principal phenomenon we initially seek to understand in these terms is *logic*. The first way logical vocabulary differs from the semantic and pragmatic metavocabularies considered so far is that it is an intrinsic, rather than an extrinsic metavocabulary for codifying reason relations. The rules by which logical vocabulary is introduced to extend any arbitrary nonlogical base vocabulary appeal to nothing more than the reason relations sentences of the base vocabulary stand in to one another.

To appreciate the second basic feature distinguishing the specifically logical codification of reason relations, one must think about the metavocabulary by which logical vocabulary is introduced. There are many vocabularies that can do this, including Hilbert-style axiomatics and Tarskian model theory. We use a version of Gentzen's sequent-calculus approach. Gentzen's basic innovation was to treat reason relations, paradigmatically implications, as objects, called "sequents," that can be referred to and manipulated, and their metainferential relations made explicit in a mathematical metavocabulary. The sequent-calculus metavocabulary can be thought of as applying to an arbitrary nonlogical base vocabulary (that is, a vocabulary whose lexicon consists of logically atomic sentences). The sequent "$\Gamma \mathbin{\vdash\mkern-10mu\sim} \Delta$" says that in the base vocabulary, the set of sentences Γ implies the set of sentences Δ. (The sequent "$\Gamma, A \mathbin{\vdash\mkern-10mu\sim}$" says that the sentence A of the base vocabulary is incompatible with the set of sentences Γ. The empty right-hand side of the turnstile says that the set of sentences on the left-hand side is materially incoherent.) Metainferential relations among sets of implications/incompatibilities of the base vocabulary, in the form of sequents, can then be expressed with horizontal lines, which are understood as saying that whenever all the sequents above the line are good (the reason relations they codify hold in the base vocabulary), then so are all the sequents below the line. In this way, structural metainferential relations among sequents can be expressed. For instance,

$$\frac{\Gamma \mathbin{\vdash\mkern-10mu\sim} \Delta}{\Gamma, A \mathbin{\vdash\mkern-10mu\sim} \Delta}$$

says that the top sequent can be weakened on the left by the sentence A, in the sense that adding A as a further premise to $\Gamma \mathrel{\vert\!\sim} \Delta$ will not turn that good implication into a bad one. If "Γ", "Δ", and "A" are read as schematic letters standing for arbitrary sets and sentences, rather than as constants standing for particular ones, this formula says that the reason relations of the base vocabulary is structurally monotonic, in the sense that adding premises never infirms implications. This sequent-calculus metavocabulary allows for efficient expression of the reason relations that hold in any arbitrary base vocabulary, including metainferential relations. But it is essentially just a notation, requiring no substantial additional conceptual resources beyond what is provided by the base vocabulary whose nonlogical implications and incompatibilities it specifies explicitly.

Perhaps surprisingly, the spare sequent-calculus notation (a version of which we have already been using informally in this *Introduction*) turns out to be sufficient to formulate rules for adding logical vocabulary to any arbitrary base vocabulary, and (most importantly) computing the reason relations of the extended vocabulary from those of the base. The idea is first to extend the lexicon of the base vocabulary, by syntactic rules that specify that the base lexicon is included in the logically extended lexicon, and that if A and B are sentences of the extended lexicon, then so are A, $A \rightarrow B$, $A \wedge B$, and $A \vee B$. The metainferential connective rules then specify the reason relations that (sets of) sentences of the logically extended lexicon stand in to one another, entirely in terms of sequents codifying the reason relations that sentences of the base lexicon stand in. These include rules such as:

$$\frac{\Gamma, A \mathrel{\vert\!\sim}}{\Gamma \mathrel{\vert\!\sim} \neg A}$$

and

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

The complete logically extended vocabulary (lexicon of logically complex sentences plus reason relations among them) can then be computed from the base vocabulary (lexicon of logically atomic sentences plus reason relations among them). We say that a corresponding logically extended vocabulary can be *elaborated from* any arbitrary base vocabulary. Implications and incompatibilities (and metainferences involving them) that hold in every logical extension of a base vocabulary, no matter what base vocabulary it is elaborated from, can then be said to hold in virtue of logic alone. These are sequents such as $\Gamma, A \rightarrow B, A \mathrel{\vert\!\sim} B$ that can be shown to hold no matter what implications hold in the base vocabulary, simply

in virtue of the rules for determining the reason relations that logically complex sentences stand in (in the context of the basic structural principle of Containment—the principle that among the conclusions implied by any premise-set are the members of that premise-set).

The sequent-calculus vocabulary is accordingly a rational meta-vocabulary—a vocabulary for specifying the reason relations of some other vocabulary—that has the special feature that it permits the elaboration of arbitrary base vocabularies into vocabularies over lexicons that extend the lexicons of the base vocabularies by adding logically complex sentences formed by combining the sentences of the base vocabulary with logical operators. Rules for those operators formulated in the sequent-calculus vocabulary conservatively extend the reason relations of the base vocabulary, in the sense that the implications and incompatibilities that hold among logically atomic sentences in the logically extended vocabulary are just those that already held among them in the base vocabulary. And the connective rules formulated in the sequent-calculus vocabulary do this while appealing to no resources outside of those provided already by the reason relations of the base vocabularies to which they apply and which they conservatively extend.

That is, sequent-calculus metavocabularies are *intrinsic* rational metavocabularies of a distinct kind. What distinguishes them from other intrinsic rational metavocabularies (examples of which will be introduced just ahead) is their serving to extend arbitrary base vocabularies by adding to them a further sort of vocabulary: logical vocabulary. And they elaborate all the reason relations of the extended vocabulary solely from the reason relations of the base vocabulary to which they are applied. This is no doubt a remarkable achievement. But one might nevertheless want to know: what is it good for? Why do this? The answer to this question provides the third, most specific characterization of what distinguishes *logical* vocabularies from all other kinds.

The idea we pursue is that logical vocabulary is distinguished above all by the characteristic *expressive* role that it plays with respect to the reason relations of the base vocabulary to which it is added. When the reason relations of the logically extended vocabulary are suitably elaborated from those of a base vocabulary, it becomes possible for the first time to *say explicitly*, in the extended vocabulary, what implications and incompatibilities hold in that base, and also in its logical extension. Logical vocabulary makes explicit the reason relations of *any* nonlogical vocabulary to which it is added and from which its own reason relations are elaborated, paradigmatically by rules expressed in a sequent-calculus metavocabulary. When introduced by suitable sequent rules, a set of sentences Γ implies the conditional $A \rightarrow B$ just in case Γ together with A implies B, and Γ implies the negation $\neg A$ just in case A is incompatible

with Γ. (This is what the sequent rules for the conditional and negation displayed above stipulate.) Logically extending a base vocabulary makes it possible to *say*, in the resulting vocabulary, what the reason relations of that base vocabulary are (and as a bonus, also those of the logically extended vocabulary elaborated from it). In addition to these proper, first class, *expressive* sentential connectives, there are also auxiliary, second class, merely *aggregative* connectives of conjunction and disjunction.

Logical vocabulary, we say, is *el*aborated from and *ex*plicative of (LX for) the reason relations of arbitrary base vocabularies. Because it can extend any and every base vocabulary (that meets the most minimal conditions), the ideal logic is *universally* LX. Logical vocabulary can be demarcated by this distinctive *explicative* role. The constellation of the sequent calculus metavocabulary and the logical vocabulary it introduces stands in an intrinsic-explicative metalinguistic relation to the reason relations of any base vocabulary whatsoever. In general, it is not hard to arrange that the reason relations of some extended language can be computed entirely from the reason relations of any arbitrary base vocabulary: that the extended reason relations can be *elaborated from* those of the base. Care is needed in formulating the rules that introduce the logical connectives, however, in order to see to it that they can perform their full *explicative* function. The rules of the logics we propose can be shown to be *expressively complete* in a strong sense. For any arbitrary set of base implications and incompatibilities, it is possible to compute an implication (or a set of implications for some of the systems below) in the logically extended vocabulary that is derivable just in case exactly those reason relations in the base vocabulary hold. And for any set of implications-and-incompatibilities in the logically extended vocabulary, there is a set of reason relations in the base vocabulary that holds just in case those logically complex reason relations are derivable.

It is even harder to formulate a set of definition rules for sentential logical connectives that is *universally* LX, in the sense that a logically extended vocabulary can be elaborated from and will then be explicative of *any and every* base vocabulary. In particular, almost all extant logics either presuppose that the base vocabularies they extend satisfy strong global structural constraints—paradigmatically the monotonicity and transitivity at the core of traditional understandings of specifically logical consequence as a kind of closure operator—or retroactively impose some such global structure, thereby failing to be conservative over some substructural base vocabularies. While we believe that purely *logical* consequence does have a global closure structure (and that logical inconsistency is monotonic), we argue that this is not in general true of *non*logical reason relations, and so is not and cannot be true of the full consequence (and incompatibility) relations of logical extensions of arbitrary base vocabularies.

The philosophy of logic that takes the ideal logical vocabulary to be one that is universally LX—that the reason relations of the logically extended vocabulary can be elaborated from and are explicative of any arbitrary base vocabulary to which the logical connective rules are applied—we call "logical expressivism." We show how to define logics that fully satisfy this ideal, and then how their capacity to express reason relations can be further augmented to include local metainferential structural properties of arbitrary base vocabularies. Our favorite (due to Kaplan (2018) and Hlobil (2018)), which we call "NMMS," for "NonMonotonic, MultiSuccedent" logic, is expressively complete for arbitrary substructural material base vocabularies, including nonmonotonic and nontransitive ones. It is supraclassical when applied to base vocabularies meeting the most minimal structural condition (Containment, which requires that every premise-set implies all the premises it explicitly contains), and it determines exactly the consequence relation of classical logic when applied to base vocabularies that consist entirely of instances of Containment. (There is also a supraintuitionist version that converges on the consequence relation of intuitionist logic under the same circumstances (Hlobil, 2016).) We call these and the various extensions of them that we define "expressivist logics."

0.4 Implication-Space Intrinsic-Explicative Metavocabularies for Reason Relations and Conceptual Roles

Intrinsic-explicative rational metavocabularies, of which logical vocabularies are a paradigm, are philosophically important because they specify the "rational forms" of the reason relations that the above-mentioned isomorphism shows are common to the richer, more committive specifications of those reason relations provided by semantic and pragmatic metavocabularies. Characterizing those common rational forms requires abstracting away from the substantive concepts *extrinsic* to the base vocabularies for which they are metavocabularies that permit semantic and pragmatic *explanations* of reason relations. If the devil is in the (extrinsic, semantic and pragmatic) details, then the angels will be found working with rational forms at the more abstract (intrinsic) level of reason relations alone. Logical vocabularies make reason relations explicit in terms that appeal only to the conceptual resources supplied by the base vocabularies from which they are conservatively elaborated. They are in that sense *intrinsic* vocabularies for specifying reason relations. Logical vocabularies, however, are not purely *meta*vocabularies, in the sense in which semantic and pragmatic rational metavocabularies are. The sequent-calculus vocabularies in which we say how to elaborate arbitrary base vocabularies into logically extended vocabularies with the capacity to codify reason relations

are genuine metavocabularies in that sense. Like semantic and pragmatic metavocabularies, they do not extend the base vocabularies for which they are metavocabularies. They are *purely* metalinguistic, talking *about* expressions in the base vocabulary, rather than *using* them. But the new, logically complex sentences introduced by adding logical vocabulary to base vocabularies by rules formulated in the sequent-calculus metavocabulary do use the expressions of the base vocabulary, rather than merely mentioning them. Logical vocabulary is a hybrid or mongrel kind of metavocabulary. It plays the expressive role of expliciting reason relations: making them explicit, constructing sentences intelligible as *saying that* relations of implication and incompatibility hold. That is a broadly metalinguistic expressive function. But logical vocabulary performs that explicative expressive function by using the sentences whose reason relations it articulates, rather than by talking *about* them (mentioning them).

These observations raise the question of whether there is a purely intrinsic-explicative vocabulary for specifying reason relations that is a rational *meta*vocabulary in the sense of being genuinely and wholly metalinguistic. The answer is "yes," and seeing how such a metavocabulary works and what it can do illuminates both the genus of intrinsic-explicative vocabularies for reason relations, which it shares with the more familiar logical species, and the concept of reason relations (and so vocabulary) as such. Our candidate, informed by work due to Daniel Kaplan (2022), is an implication-space metavocabulary for specifying both reason relations and the conceptual roles sentences play in virtue of standing to one another in such reason relations. Very roughly, where Gentzen's sequent-calculus metavocabulary treats implications as basic objects in a proof-theoretic formalism, Kaplan's implication-space metavocabulary treats them as basic objects in a model-theoretic formalism. It represents the current state of the art in inferentialist semantics.

Inferentialists have long thought that the universe from which semantic interpret*ants* are drawn or from which those interpretants are built—the analogue of the universe of mereologically structured worldly states out of which semantic interpretants (propositions) as pairs of sets of truth-making states and falsifying states are built—should consist of implications (including incompatibilities coded as implications) and sets of them. Until now, however, inferentialists had assumed that the items that are in the first instance semantically interpret*ed* should be the same declarative sentences that stand to one another in reason relations of implication and incompatibility. This methodological commitment was thought to be controversial only from below, inasmuch as some representationalists in the model-theoretic tradition begin instead with singular terms and predicates, building sentences out of subsentential parts, thought of as more basic. By contrast, inferentialists understand such components as emerging only as

the result of analysis of sentences by assimilating sentences into equivalence classes according to which lexical-syntactic substitution relations among them preserve the goodness of implications among them (see Brandom, 1994, Chapter Six).

Kaplan's (2022) first conceptual innovation was the idea that thoroughgoing inferentialists ought to treat the most basic units being interpreted, no less than the semantic intepretants assigned to them, as being implications, rather than the sentences that make up their premises and conclusions. Only at a second, subsequent stage, would semantic interpretation be extended from implications to the sentences they contain. He accordingly begins with a universe of candidate implications, together with a partition of that universe into a distinguished set of *good* implications—ones whose conclusions really follow from their premises—and the rest. This universe of candidate implications with a distinguished subset is an *implication space*. Given a lexicon of sentences, it is just the space consisting of all the ordered pairs of sets of sentences of that lexicon. Thought of as the space from which semantic interpret*ants* are drawn, such an implication space is analogous to Fine's space of states, partitioned into a distinguished set of possible states, and a residue of impossible states. Any base vocabulary determines such an implication space, since the lexicon of the vocabulary suffices to define the points (candidate implications as ordered pairs of sets of sentences of the lexicon), and the reason relations of the vocabulary suffice to determine the distinguished set of good implications. Since that is all that implication-space semantics requires to define (at the second stage of semantic interpretation) the implicational roles played by sentences, it will be an intrinsic semantics, appealing only to the conceptual resources afforded by the base vocabularies for which it offers a semantic interpretation.

The new idea is to begin the process that will end in semantic interpretation of sentences by using the points of the implication space generated by any arbitrary base vocabulary—the candidate implications, both good and bad—also as the source of the items to be semantically interpreted. What could count as a semantic interpretation of an implication? What would make one think the idea even makes sense? Here is a rationale. We are exploring the idea of understanding meaning to begin with in terms of *reasons* instead of understanding it in terms of *truth*. That is to understand meaning in terms of a dyadic *relation* (between sets of sentences) instead of in terms of a monadic *property* (of sentences). On the approach that takes truth as basic, one starts with assignments to sentences of a truth *value*: as true or false, correct or incorrect, good or bad (*as* a representation). However, although assignments of truth values are the beginning of semantic interpretation on this approach, they are not the end. To get a notion of meaning that corresponds to what one grasps (however

imperfectly) when one understands a sentence, one must advance from consideration of truth *values* to consideration of truth *conditions*. (One must add to a semantic conception of the Fregean *Bedeutung* of a sentence a semantic conception of its Fregean *Sinn*.)

Thought of as a universe of potentially semantically evaluable items, the implication space determined by any given base vocabulary has a natural analogue to truth values, but on the side of reasons. For reason relations of implication and incompatibility (represented as pairs of sets of sentences of the lexicon) are divided into good ones and bad ones, ones that hold and ones that do not, correct implications (codifying real reasons) and incorrect ones (codifying possible but not actual reasons). At the extensional semantic ground level, one can say *that* a sentence is true, and in the reason-based setting one correspondingly can say at the extensional semantic ground level *that* an implication is good or an incompatibility holds. Given that analogy, the question becomes: what stands to implication (reason relation) *values* (good/not-good) as truth *conditions* stand to truth *values*?

The idea behind truth conditions (and Fine's generalization to truth-makers and falsifiers) is that apart from the question of whether a truth-candidate actually *is* true or false, there is the question of what it would take to *make* it true—what things would have to be like for it to count as correct in this distinctive semantic sense. The idea behind the first stage of implication-space semantics is that apart from the question of whether a candidate implication actually *is* good (according to the partition of the space of candidate implications into good and bad determined by the underlying base vocabulary), there is the question of what it would take to *make* it good. In the special case of reason relations that already do hold, candidate implications that are good, this takes the form of asking about the circumstances under which it would *remain* good. That is the *range of subjunctive robustness* of the implication.

To make this thought slightly more definite, we can consider a single-succedent sequent $\Gamma \mathrel{|\!\sim} A$. Suppose it is a good implication—that A actually follows from Γ according to the reason relations of the base vocabulary. We can think of the range of subjunctive robustness of this implication as the set of additional premises that could be added to Γ without turning it from a good to a bad implication. The range of subjunctive robustness of $\Gamma \mathrel{|\!\sim} A$ would then consist of all the good implications of the form $\Gamma, \Delta \mathrel{|\!\sim} A$. If $\Theta \mathrel{|\!\sim} A$ does *not* hold (is not in the distinguished set of good candidate implications of that particular implication space) then its range of subjunctive robustness is the set of all good implications that result from adding further premises that would make it good: all the $\Theta, \Delta \mathrel{|\!\sim} A$ that *do* hold. Those sets represent what additional premises are required to make the original candidate implication into one that is good. The range of

subjunctive robustness of a candidate implication is its semantic intepretant in the form of its *good-makers*, as in Fine's truth-based semantic setting the semantic interpretants of sentences are their *truth-makers* (and falsifiers).

Grasping ranges of subjunctive robustness in this sense is an essential part of understanding reason relations in ordinary vocabularies. Working with ordinary, defeasible nonlogical implications, one might think that the conclusion:

C)   The lioness will attack the gazelle.

does not follow from

P1)   The lioness can see the nearby wounded gazelle.

but that it does follow from

P2)   The lioness is healthy.
        The lioness is hungry.
        The lioness can see the nearby wounded gazelle.

That is understanding something about what it would take to make $P1 \mathop{\vdash\!\sim} C$ a good implication, even though as it stands it is not. $P2$ is adding premises that yield a good implication. And if one thinks further that C would follow from

P3)   The lioness is healthy.
        The lioness is hungry.
        The lioness can see the nearby wounded gazelle.
        Today is Tuesday.

but would not follow from

P4)   The lioness is healthy.
        The lioness is hungry.
        The lioness can see the nearby wounded gazelle.
        The lioness will immediately be struck by lightning.

then one appreciates something important about the range of subjunctive robustness of the implication $P2 \mathop{\vdash\!\sim} C$.

The ranges of subjunctive robustness of candidate implications are their "goodness" conditions, as truth conditions are the "goodness" conditions of sentences. For an implication to be good in the reasons-first semantic setting is for its premises to provide reasons for its conclusion

(or reasons against, in the case of incompatibilities), while for a sentence to be good in the truth-first semantic setting is for it to be true. The advance from a conception of semantic goodness to a conception of meaning is the advance to consideration of circumstances under which a reason relation or sentence *would* be good. In the truth-maker/falsifier semantic setting, the circumstances are sets of worldly states, specified in some vocabulary that is not in general part of the base vocabulary whose sentences are being semantically interpreted. In the implication-space semantic setting, the circumstances are additional premises (and, in the fully general multisuccedent case also additional conclusions) that would make or keep the reason relation good. By contrast to the truth-maker setting, in the implication-space setting, those further premises and conclusions are just more sentences of the lexicon of the base vocabulary. That is why implication-space semantics counts as *intrinsic*.

Nothing has yet been said about why or how implication-space constructions can be *explicative* of the conceptual contents conferred on sentences in the lexicon by the reason relations of the base vocabulary, from which the implication space is derived. That expressive achievement is the task of the second stage of semantic interpretation in the implication-space setting. During that second stage, semantic interpretants for sentences (in the form of conceptual roles) are computed from the semantic interpretants of implications (their ranges of subjunctive robustness) in which those sentences appear as premises or conclusions. Essentially, the conceptual role of a sentence is represented as the ordered pair of the set of items with the same range of subjunctive robustness as the bare occurrence of the sentence as a premise and the set of items with the same range of subjunctive robustness as the bare occurrence of the sentence as a conclusion. It turns out that conceptual roles of this form perspicuously and tractably codify the contributions the occurrence of a sentence makes to the goodness of implications and incompatibilities in which it occurs as premise or conclusion.

One measure of the extent to which the implication-space specification of the conceptual roles sentences play in virtue of standing to other sentences in the reason relations they do is an expressively powerful one is that various semantically significant relations among sentences correspond to natural operations on conceptual roles specified in these terms. To begin with, the sequent-calculus rules that extend arbitrary base vocabularies by adding sentential logical connectives all correspond to simple ways of combining implication-space conceptual roles. In this way, a model-theoretic inferentialist semantics becomes available that is sound and complete for the aforementioned expressive logic NMMS. The implication-space semantics shows how to compute the conceptual roles of arbitrary logically complex sentences from the conceptual roles

of logically atomic sentences of any base vocabulary—even when the base vocabulary, and so its (conservative) logical extension, are radically substructural, including those that do not satisfy the metainferential structural closure conditions of monotonicity and transitivity. To do this, the implication-space rational metavocabulary must make explicit the conceptual roles played by sentences of all those base vocabularies, as well as their logical extensions. It is universally explicative of sentential conceptual roles. And since implication spaces can be constructed using no resources other than those supplied by the spare specifications of arbitrary, even substructural base vocabularies—just sets of sentences and set-theoretic constructions from them representing their reason relations—the implication-space model-theoretic semantics qualifies as a *universal* intrinsic-explicative rational metavocabulary.

One reason to think that the implication-space metavocabulary for explicating the conceptual roles sentences play in virtue of standing to one another in the reason relations that define any base vocabulary cuts at important joints is that we can erect on top of it a calculus codifying metainferential relations among those conceptual roles (see also Kaplan, 2022). Metainferences of various kinds can be defined precisely, systematic combinations of them recursively constructed, and the effects of those combinations computed. The result is a principled botanization of constellations of metainferences that offers revealing characterizations of a number of logics that have been the subject of intense interest among logicians and philosophers of logic over the past few decades. The trilogics, paracomplete K3 (Strong Kleene) and its dual, paraconsistent LP (Logic of Paradox), for instance, show up as the logics of premissory and conclusory metainferences, respectively—in virtue of the same features that make them show up as logics of truth-value gaps and gluts, respectively, in the truth-based semantic setting. The nontransitive classical logic ST (Strict-Tolerant) and the closely related logic TS (Tolerant-Strict) fall into place neatly in relation to them. In treating metainferential relations among conceptual roles as objects that can be combined and manipulated, this calculus stands to conceptual roles as the sequent calculus stands to the sentences that are the relata of the implication relations it codifies as sequents. This intrinsic rational metametavocabulary, built on top of the implication-space inferentialist model-theoretic semantics for conceptual roles, provides the expressive power to make explicit a hitherto unexplored level of metainferential reason relations among those roles, and thereby offers an illuminating new semantic perspective on the relations among a variety of well-studied logics.

The implication-space metavocabulary provides a model-theoretic semantics for the conceptual roles sentences play in virtue of standing to one another in reason relations of implication and incompatibility.

It is a reason-based inferentialist semantics, rather than a truth-based representational semantics like truth-maker semantics. By contrast to the proof-theoretic treatment of reason relations by the sequent calculus, the implication-space metavocabulary assigns sets of implications as the semantic interpretants of implications, and set-theoretic constructions out of those sets as the semantic interpretants of sentences, and then operates on and manipulates those semantic interpretants to codify reason relations and conceptual roles. In fact, it does so in a way that can be shown to be isomorphic to the truth-maker model-theoretic semantics. Each constructs bilateral semantic interpretants for sentences (truth-makers and falsifiers in the one case and premissory and conclusory roles in the other) from a universe consisting of a set of points (states and candidate implications, respectively) with a distinguished subset (possible states and good implications). In both cases, the universe is taken to be structured by a commutative monoid (fusion of states and a corresponding operation combining candidate implications according to their ranges of subjunctive robustness). Nonetheless, the implication-space metavocabulary provides an intrinsic semantics, since it appeals to nothing that is not made available by the base vocabulary to which it is applied: sets of sentences and their reason relations. Implication-space semantics is something like the *intrinsification* of truth-maker semantics—in a way formally analogous to, but more expressively powerful than, Fine's use of intrinsic "canonical models."

When this structural isomorphism of implication-space and truth-maker semantics—which holds between the universes from which semantic interpretants are drawn, the interpretants themselves, and the way reason relations of consequence and incompatibility are determined for sentences in terms of their semantic interpretants—is appreciated in detail, and considered in context with the orthogonal isomorphism at the level of reason relations between the truth-maker alethic modal semantic metavocabulary and the deontic normative bilateral pragmatic metavocabulary, it becomes clear that the implication-space semantics makes explicit the abstract rational forms common to those two extrinsic-explanatory metavocabularies of meaning and use. Those rational forms are just the conceptual roles the implication-space semantics characterizes. As an intrinsic-explicative rational metavocabulary, the implication-space explication of those conceptual roles abstracts away from the extrinsic appeals to the possibility or impossibility of fusions of worldly states and the propriety or impropriety of constellations of practical commitments to accept and reject that the truth-maker semantic and bilateral pragmatic metavocabularies call on to explain implication and incompatibility relations. As the pure (intrinsic) metavocabulary of conceptual roles, inferentialist implication-space semantics is our candidate for the language of the angels, who think by manipulating the abstract rational forms

shared by representations of worldly states on the objective side of the intentional nexus and manifestations of normative practical attitudes on the subjective side. We are pleased and grateful to report that the intrinsic-explicative metavocabularies of implication-space semantics and logic can also be used by embodied fallen creatures to achieve that angelic rational self-consciousness.

0.5 Conclusion

The topic of this book is metavocabularies of reason, or rational metavocabularies. "Vocabulary" is a term of art. We mean by it a set of items called "sentences," (a "lexicon") and a set of relations on (sets of) those sentences that we call "reason relations." The justification for calling the elements of the lexicon "sentences" is that they stand to one another in reason relations. So the question becomes: what makes a set of relations deserve to count as "reason" relations in a sense that in turn justifies calling what stands in such relations "sentences"?

   As we use the term, there are two principal kinds of reason relation: "implications" and "incompatibilities." Those indeed sound like varieties of rational relation—but clearly the issue then becomes what qualifies relations as those of implication or consequence (a kind of rational inclusion) and incompatibility or inconsistency (a kind of rational exclusion). For the last century or so, logicians have offered responses to this question that look to the *structure* of the relations in question. Inspired by structural features of the best-behaved logics of their day, Tarski and Gentzen (founders of the model-theoretic and proof-theoretic metavocabularies for logic) agreed that consequence relations are reflexive, monotonic, and transitive, incompatibility relations are monotonic, and that consequence and incompatibility are related by the implicational explosion of inconsistent sets of sentences: inconsistent premise-sets imply every sentence. For both principled and empirical reasons, we deny that *any* of these global structural constraints should be taken to characterize reason relations in general, as we understand our topic. So we cannot appeal to this venerable idea to demarcate reason relations. How else, then?

   Our answer, in short, is that reason relations are the common topic of a specific constellation of *meta*vocabularies. These are *rational* metavocabularies, in that the reason relations among their sentences depend, in various ways characteristic of the kind of rational metavocabulary they are, on the reason relations of the base vocabularies for which they are metavocabularies. Vocabularies whose reason relations do not depend on the reason relations of any other vocabularies are not rational metavocabularies in this sense—though to be vocabularies at all they must consist of sentences standing in reason relations. The ultimate goal is to understand the sentences of these base vocabularies—not just the reason relations among

them, but also their *use* to make claims and the *meanings* they express concerning how things are with the subject-matter they make it possible to talk about. We aim to illuminate those important topics by considering the reason relations the sentences stand in to one another, that is, the vocabulary that comprises those sentences. And the strategy is to understand base vocabularies—their use, the meanings of their expressions, and, as our central focus, the *reason relations* among sentences and the *conceptual roles* the sentences play in virtue of standing in those reason relations—in terms of *meta*vocabularies that have the expressive power to specify those uses, meanings, reason relations, and conceptual roles.

This strategy for specifying this new topic of discursive reason relations, and the notion of a *vocabulary* defined in terms of it, is *metalinguistic functionalism*. We understand reason relations in terms of the relations that articulate a whole constellation of *rational metavocabularies*: vocabularies for specifying reason relations. Reason relations, we want to say, are the common topic of just this constellation of metavocabularies. Our selection of the metavocabularies to consider is principled, and the space in which their relations to one another show up has a clear structure.

They are not, however, all created equal. We will take a frankly inegalitarian attitude toward them. The base vocabularies it is most important to understand are those codifying the reason relations of autonomous discursive practices—language-games one could play though one played no other, discursive practices that could be engaged in by practitioners who engaged in no others that were not fragments of them. Paradigmatically, these are natural languages. Picking them out in terms of their practical autonomy is demarcating them in a *pragmatic* metalanguage. Any such autonomous discursive practices must confer on some performances the significance of claimings, sayings in the sense of assertings. We think that means that they must also contain *reasoning* practices, in the form of claimings that serve as challenging other claimings by offering reasons against them, and other claimings that serve as defending them, by offering reasons for them. The relations between *reasoning*, as something done, and the reason *relations* of incompatibility (structuring reasons against) and implication (structuring reasons for) that are our proximal targets are complex. But the case for calling them *reason* relations ultimately turns on their relation to *practices* of rationally challenging and defending claims. Pragmatic metavocabularies, which specify reason relations in terms of the use of sentences to make claims and defend and challenge them, accordingly occupy a privileged place among rational metavocabularies. The substantive account they offer of what reason relations are is where we begin.

Representational semantic metavocabularies in the model-theoretic tradition offer a very different kind of substantive account of consequence

and incompatibility. This *Introduction* began with the thought that one might usefully illuminate the crucial relations between use and meaning (each the subject of an important philosophical tradition) by better understanding the relations between what is expressed by pragmatic and semantic metalanguages. Constructing, as we do, an isomorphism between suitably chosen versions of these kinds of metavocabulary shows that they indeed have a common topic. The content of identity claims, Frege taught us, is that two different senses pick out a common referent. Isomorphisms are equivalence relations, and the process of treating an equivalence as an identity is the essence of abstraction. The above-mentioned isomorphism accordingly shows that bilateral normative pragmatic metavocabularies and truth-maker semantic metavocabularies indeed have a common abstract topic: relations of consequence and incompatibility. That is what we mean by "reason relations."

Identifying this common target for accounts couched in these two principal kinds of extrinsic-explanatory metavocabulary brings into relief the possibility of another genus of rational metavocabularies addressing the same subject-matter. Intrinsic-explicative rational metavocabularies confine themselves to the conceptual resources supplied by the base vocabularies they address, and seek merely to express, rather than to explain, their reason relations. Logical vocabularies make explicit reason relations of implication and incompatibility in any arbitrary base vocabulary by extending it with conditional and negated sentences formed ultimately from the lexicon of that base vocabulary and then computing the reason relations of the extended vocabulary from those of the base. The implication-space rational semantic metavocabulary provides the expressive power to construct and manipulate the conceptual roles sentences play in virtue of standing to one another in the same reason relations articulated by logical vocabulary. These two principal kinds of intrinsic-explicative rational metavocabulary are related to each other not by the sort of isomorphism that relates the (right kind of) semantic and pragmatic extrinsic-explanatory rational metavocabularies. They are related to one another rather by the demonstrated soundness and completeness of the implication-space semantics for the most flexible and expressively powerful logic.

The internal structure of this constellation of two families comprising four kinds of rational metavocabulary is completed by showing how to construct an implication-space model corresponding to any truth-maker model of an arbitrary base vocabulary, by a mapping that preserves reason relations and conceptual roles. Doing that concludes the argument that it is the same thing that is explicated (each in its own way) by logical and implication-space metavocabularies, and that is explained (each in its own

way) by semantic and pragmatic metavocabularies: the reason relations and conceptual roles of sentences.

The illumination we believe is offered by telling the story we do here is of two kinds. The more general is the precise metalinguistic functionalist identification of a philosophical subject-matter—namely reason relations and the conceptual contents they articulate—as the common topic of metavocabularies of these four kinds, related to one another as we have shown they can be. The more specific comprises the particular versions of each of these kinds of metavocabulary that we elaborate: the more fine-grained bilateral deontic normative pragmatic metavocabulary, the specific consequence relation and way of relaxing global structural constraints in truth-maker semantic metavocabularies, the nonmonotonic multisuccedent logic that is provably expressively complete when extending even radically open-structured base vocabularies, and the implication-space semantic metavocabulary for conceptual roles. The details of those four particular ways of working out general conceptions of the expressive tasks characteristic of the different kinds of metavocabulary are essential to showing that they can stand in the relations to one another that reveal and delineate rational forms as their common topic.

The illumination an account offers of the phenomena it addresses crucially depends on its achieving the right level of abstraction. It must be general enough to provide a satisfying unification of otherwise disparate phenomena while specific enough to engage convincingly with the distinguishing details of the items it brings together. The metalinguistic functionalist picture that is elaborated in the body of this book is our candidate for achieving such a balance. The phrase "level of abstraction" is a loose way of talking about a structured set of isomorphisms. That is exactly how we have articulated the functional system of rational metavocabularies.

Every kind of rational metavocabulary is a form of rational self-consciousness. They are characteristic ways of talking and thinking about the rational structure of talking and thinking. For each is a distinctive way of expressing the reason relations that articulate the conceptual contents sentences acquire by being used as they are and standing to worldly states in the representational relations they do. Laying out the fundamental relations among these different kinds of rational metavocabulary is accordingly unpacking the internal structure of rational self-consciousness as such.

Notes

1 "Thought" in the sense of "thinkable," not in the sense of episodic "thinkings" of it (Frege, 1956).
2 A closely related result was first proven in (Hlobil, 2022a).