

Chapter 1: Scrutability and the *Aufbau*

David J. Chalmers

1 Primitive Concepts

It is a familiar idea in philosophy that many concepts can be wholly analyzed as structures made up of simpler concepts. For example, Aristotle held that *man* can be defined as *rational animal*. This suggests that the concept *man* is a complex concept built out of the simpler concepts *rational* and *animal*.

Many philosophers go further, suggesting that all concepts are built up from a limited supply of primitive concepts. An idea like this is suggested by John Locke:

For all our complex ideas are ultimately resolvable into simple ideas, of which they are compounded and originally made up, though perhaps their immediate ingredients, as I may so say, are also complex ideas.

One also finds the idea in some parts of contemporary cognitive science. The linguist Anna Wierzbicka, for example, has argued that every expression in every human language can be analyzed in terms of a limited number of “semantic primes” that occur in every language. In her 1972 book *Semantic Primitives*, Wierzbicka proposed 14 semantic primes, but since then the list has expanded to 62. Her list of 62 primes is given below.

Substantives: *I, you, someone, people, something/thing, body.*

Determiners: *this, the same, other.*

Quantifiers: *one, two, some, all, many.*

Evaluators: *good, bad.*

Descriptors: *big, small.*

Mental predicates: *think, know, want, feel, see, hear.*

Speech: *say, words, true.*

Actions and events: *do, happen, move, touch.*

Existence and possession: *there is, have.*

Life and death: *live, die.*

Time: *when/time, now, before, after, a long time, a short time, for some time, moment.*

Space: *where/place, here, above, below; far, near, side, inside, touching.*

Logic: *not, maybe, can, because, if.*

Augmentors: *very, more.*

Wierzbicka's methods have been used to analyze an extraordinary range of expressions in many different languages. To give the flavor of the project, a sample analysis runs as follows.

X *lied* to Y =

X said something to person Y;

X knew it was not true;

X said it because X wanted Y to think it was true;

people think it is bad if someone does this.

Within philosophy, this idea has been developed most systematically by Bertrand Russell and Rudolf Carnap. Russell suggested that all concepts are composed from concepts of objects and properties with which we are directly acquainted. For Russell, these concepts included concepts of sense-data and certain universals, and at certain points in his writings, a concept of oneself. All other concepts were to be analyzed as constructions out of these concepts. For example, concepts of other people and of objects in the external world were to be analyzed as descriptions built up from these basic elements.

In the *Der Logische Aufbau der Welt*, Carnap pushed the project of analysis to its limit. Carnap argued that all concepts can be constructed from a single primitive concept, along with logical concepts. This concept is a concept of a certain relation between total experiences (roughly, momentary slices of a stream of consciousness) had by a subject at different times. The relation is that of phenomenal similarity in some respect. For example, if a subject has two experiences both involving a certain shade of red, they will stand in this relation of similarity. Using this simple concept, Carnap gives explicit constructions of many other concepts applying to experiences. For example, concepts of specific sensory qualities, such as that of a certain shade of red, are defined in terms of chains or circles of similarity between experiences.

Carnap went on to indicate how these concepts might be used to build up all of our concepts that apply to the external world. Spatial and temporal concepts are defined in terms of sensory

qualities. Properties of external bodies are defined in terms of spatial and temporal properties. Behavior is defined in terms of the motion of bodies. Mental states of other people are defined in terms of behavior. Cultural notions are defined in terms of these mental states and behavior. And so on.¹

Carnap's project, like most of the other projects above, is committed to what we can call a *Definability* thesis. Like the other theses I discuss in this chapter, this thesis is cast in terms of expressions (linguistic items such as words) rather than in terms of concepts (mental or abstract items) for concreteness.

Definability: There is a compact class of primitive expressions such that all expressions are definable in terms of that class.

I will say more about compactness later, but for now it suffices to think of this as requiring a small class of expressions. For most of the *Aufbau*, the class of primitive expressions included an expression for phenomenal similarity and logical expressions ('not', 'and', 'exists', and the like). Late in the *Aufbau*, Carnap went on to argue that phenomenal similarity is itself dispensable: it can be itself be defined in logical terms. If so, then primitive expressions can be restricted to logical expressions, and all other expressions can be defined in terms of these. Of course the general program of definability is not committed to as strong a claim as this.

We can say that an expression E is definable in terms of a class of expressions C if there is an adequate definition statement with E on the left-hand side and only expressions in C on the right hand side. Definition statements are required to have a certain logical form, depending on the grammatical category of E . Various different logical forms might be proposed, but the differences will not matter for our purposes. As an example, definition statements for singular terms, general terms, and predicates might be required to specify the extension of E (roughly, the entity or entities in the world that E applies to) in a form akin to the following: 'For all x , x is Hesperus if and only if x is the brightest object visible in the evening sky'; 'For all x , x is a bachelor if and only if x is an unmarried man'. If such definition statements are adequate, then 'Hesperus' is definable in terms of 'brightest', 'evening', and so on, and 'bachelor' is definable in terms of 'unmarried', 'man', and so on.

What is it for a definition statement to be adequate? Here, there are various possible criteria. Certainly one should require at least *extensional* adequacy: that is, definitions of the sort above

¹It must be acknowledged that the details are sometimes sketchy. See the start of chapter 6 for an illustration of Carnap's treatment of culture.

must be true, so that the extensions of the relevant expressions on the left and right sides are the same. But typically more is required. Suppose that as it happens, all bachelors in our world are untidy men and vice versa. Then 'For all x, x is a bachelor if and only if x is an untidy man' is true, and the definition statement is extensionally adequate. Still, this statement does not seem to give an adequate definition of 'bachelor'.

To handle these cases, it is common to require some form of stronger-than-extensional, or *intensional*, adequacy for a definition. For example, it is often required that a definition statement be analytic (true in virtue of meaning), a priori (knowable without justification from experience), and/or necessary (true in all possible worlds). A definition of 'bachelor' in terms of 'untidy man' does not meet these conditions: one cannot know a priori that all bachelors are untidy, it is not true in all possible worlds that all bachelors are untidy, and this is not plausibly true in virtue of meaning. But it is at least arguable that a definition of 'bachelor' in terms of 'unmarried man' meets these conditions.

A surprising and often-overlooked feature of the *Aufbau* is that Carnap there requires only that definitions be extensionally adequate. It is questionable whether definitions as weak as this can fulfil the various epistemological and semantic goals of the *Aufbau*. For example, while an analysis of 'bachelor' as 'unmarried man' might shed some light on the meaning of 'bachelor' and on how we come to know truths about bachelors, the same does not seem true of a definition as 'untidy man', even if that definition is extensionally adequate. In the preface to the second edition of the *Aufbau*, Carnap says that this is the greatest mistake in the project, and says that definitions should be held to a stronger, intensional, criterion of adequacy. Certainly much of the *Aufbau* can be read with a stronger criterion of adequacy in mind.

The stronger criteria of analyticity, apriority, and necessity ensure that an expression and its definition are connected semantically (that is, in the realm of meaning), epistemologically (in the realm of knowledge), and modally (in the realm of necessity and possibility). Further potential criteria include psychological criteria, to the effect that a definition somehow reflects the psychological processes involved in understanding and using an expression; formal criteria, to the effect that definitions have a certain limited complexity; conceptual criteria, to the effect that the expressions used in the definition express concepts that are more basic (in some relevant sense) than the concept expressed by the original expression; and so on.

If the definition of 'bachelor' in terms of 'unmarried man' is extensionally adequate, then truths such as 'John is a bachelor' will be logically entailed by truths such as 'John is an unmarried man' along with the definition. More generally, given certain assumptions about the language,²

then if this definition is adequate, any statement containing ‘bachelor’ will be logically entailed by a corresponding sentence containing ‘unmarried man’ in place of ‘bachelor’ (with the rest of the sentence as before), along with the definition. Given these assumptions, the Definability Thesis leads to the following thesis:

Definitional Scrutability: There is a compact class of truths such that all truths are definitionally scrutable from that class.

Here, a truth is a true sentence.³ A compact class of truths, to a first approximation, is a class of truths involving only a small class of expressions. A sentence *S* is definitionally scrutable from (or definitionally entailed by) a class of sentences *C* if *S* can be logically derived from some members of *C* and some adequate definition sentences. For example, given the relevant assumptions, sentences involving ‘bachelor’ are definitionally scrutable from sentences involving ‘unmarried man’. If we repeat this process for every definable expression, we can eventually translate every sentence of the language into a sentence in the primitive vocabulary, and the original statement will be entailed by the transformed sentence conjoined with a number of definitions.

On the *Aufbau* view, all truths are definitionally scrutable from a class of truths about the phenomenal similarity relation. In fact, Carnap holds that there is a single *world-sentence* *D*, such that all truths are definitionally entailed by the world-sentence. The world-sentence says there exist such-and-such entities are related in such-and-such fashion by the phenomenal similarity relation. If there were just three total experiences in the world, then the world-sentence would be something like: $\exists x, y, z(Rxy \& Ryz \& \neg Rxz \& \forall w(w = x \vee w = y \vee w = z) \& \neg(x = y) \& \neg(y = z) \& \neg(x = z))$, where

²We can assume that every natural-language sentence has a *regimentation* into an equivalent sentence with a clarified logical form. One can then apply definitional and logical machinery to regimented sentences in the first instance, and derivatively to unregimented sentences. If definitions are required only to be extensionally adequate, it suffices to assume that the language and the logic are extensional: that is, the logic allows one to substitute coextensive expressions (given a statement of coextensiveness), and the language is such that this substitution will not change truth-value. If definitions are required to be intensionally adequate, it suffices to assume that the language and the logic are intensional to the same degree (definitions will then need to contain a statement of cointensiveness, such as ‘Necessarily, bachelors are unmarried men’). The language may also be hyperintensional, so that cointensive expressions are not intersubstitutable in certain contexts, as long as these contexts can themselves be defined in an extensionally/intensionally adequate way (‘For all *S*, *S* believes that such-and-such iff ...’).

³The choice of sentences rather than propositions here is discussed in 2.2. A subtlety here is that not all sentences are true or false independent of context. For example, there may be no context-independent fact of the matter about whether a sentence such as ‘I am hungry’ or ‘John is tall’ is true. Where context-dependent sentences are concerned (E3), we can talk instead of the scrutability of sentences in contexts.

R is a term for the phenomenal similarity relation. If there are more total experiences than this, then there will be a longer world sentence, specifying the similarity relations that do and do not hold among the total experiences.

According to Carnap's stronger view late in the *Aufbau*, the previous world sentence D is definitionally entailed by an even more austere world sentence D' , using purely logical vocabulary. To get from D to D' , Carnap defines away the single nonlogical vocabulary item R as that relation that makes the previous world-sentence D true.⁴ If this is correct, then the highly austere truth D' definitionally entails all truths.

If we require that adequate definitions are a priori (knowable independently of experience), as is common, then Definitional Scrutability entails the following thesis:

A Priori Scrutability: There is a compact class of truths such that all truths are a priori scrutable from truths in that class.⁵

We can define a priori scrutability in parallel to definitional entailment: a sentence S is a priori scrutable from (or a priori entailed by) sentences in a class C iff S can be logically derived from some members of C along with some a priori truths. Given weak assumptions,⁶ the right-hand side is equivalent to the claim that there is a conjunction D of sentences in C such that the material conditional 'If D , then S ' is a priori.

One can characterize Analytic and Necessary Scrutability theses in a parallel way. If we require that adequate definitions are analytic or necessary respectively, then these theses will follow from Definitional Scrutability.

It is theses such as A Priori and Analytic Scrutability that give the definitional program its epistemological bite. To a first approximation, these theses suggest that knowledge of the base truths about the world might serve as a basis for knowledge of all truths about the world.

⁴That is, R is defined as that relation R' such that: $\exists x, y, z : (R'xy \& R'yz \& \neg R'xz \& \forall w (w = x \vee w = y \vee w = z) \& \neg (x = y) \& \neg (y = z) \& \neg (x = z))$. Then the new world sentence will be the resulting of replacing R in the world-sentence above by this definition. Or more straightforwardly, the world-sentence can simply say $\exists R', x, y, z : (R'xy \& R'yz \& \neg R'xz \& \forall w (w = x \vee w = y \vee w = z) \& \neg (x = y) \& \neg (y = z) \& \neg (x = z))$.

⁵A more elaborate definition of a priori scrutability is given in 2.5, and a more elaborate discussion of what it is for a sentence to be a priori is in 4.1.

⁶In one direction, it suffices to assume that all conjunctions are logically derivable from their conjuncts (this is trivial in the finite case, but slightly less trivial if infinite conjunctions are allowed, as may be necessary for some purposes). In the other direction, it suffices to assume that when B is logically derivable from a set A of premises, a conditional 'If D then B ' is a priori, where D is a conjunction of the premises in A , and that a priori conjuncts can be detached from the antecedents of a priori conditionals without loss of apriority.

To make this vivid: suppose that Laplace's demon is given all the base truths about our world. Given Definitional Scrutability, then as long as the demon knows all the definitions and can engage in arbitrary logical reasoning, then the demon will be able to deduce all truths about the world. Given A Priori Scrutability, then as long as the demon can engage in arbitrary a priori reasoning, then it will be able to deduce all truths about the world. For example, if Carnap is right, then the demon should be able to derive all truths about the world from a world sentence such as D or D' .

2 Objections to the *Aufbau*

The *Aufbau* is widely held to be a failure. It is widely held that no project like it can succeed. These doubts have a number of sources. Perhaps the best known problems for the *Aufbau* are arguments that Carnap's primitive vocabulary cannot do the work it needs to do. Two of these are specific criticisms of Carnap's constructions from phenomenal vocabulary, while another two are general criticisms of constructions from phenomenal vocabulary or from logical vocabulary.

First: In *The Structure of Appearance* (1951), Nelson Goodman argued that Carnap's definition of sensory qualities in terms of the primitive of recollected phenomenal similarity is unsuccessful, as there can be circles of similarity among total experiences that do not correspond to a single sensory quality. One problem raised by Goodman is that of "imperfect community": a similarity circle can satisfy Carnap's definition of a sensory quality even when some members of the circle share one quality (phenomenal redness, say) and others share another quality (phenomenal blueness). Another problem is that of "companionship": if two distinct qualities always occur together in total experiences, Carnap's definition will not distinguish them.

Second: In "Two Dogmas of Empiricism" (1951), W.V. Quine argued that Carnap's definition of space-time points in terms of the phenomenal field is unsuccessful, as it requires nonphenomenal notions that violate his own criteria of adequacy. Carnap defined "Quality q is at x, y, z, t " by specifying certain principles for assigning qualities to spacetime points that must be obeyed as well as possible, but this does not yield a definition that can be cast entirely in terms of phenomenal notions and logic.

Third: In "The Problem of Empiricism" (1948), Roderick Chisholm argued that no definitional analysis of statements about the external world in purely phenomenal terms can succeed. 'There is a doorknob in front of me' (P) must be analyzed as a complex conditional along the lines of 'If I had certain experiences, certain other experiences would follow' (R): for example, 'If I experience a certain sort of attempt to grasp, I would experience a certain sort of contact'. Chisholm argues

that no such R is entailed by P , as one can always find a further sentence S (e.g. specifying that one is paralyzed and subject to certain sorts of delusions of grasping that are never accompanied by experiences of contact) that is consistent with P such that $S \& P$ entails $\neg R$.

Fourth: In “Mr. Russell’s Causal Theory of Perception” (1928), the mathematician Max Newman pointed out a general problem for the more ambitious project of reducing the primitive vocabulary to logical structure alone. The problem was pointed out simultaneously by Carnap himself late in the *Aufbau*.⁷ Given a purely logical vocabulary, the ultimate world-sentence (like D' above) will specify simply that there exist certain objects, properties, and relations that stand in certain patterns of instantiation and coinstantiation. Newman and Carnap observe that as long as we are liberal enough about what we count as a property or a relation, then this world-sentence will be satisfied almost vacuously.⁸ Carnap responds by suggesting that the properties and relations in question must be restricted to “natural” (or “founded”, or “experientable”) properties and relations. This requires an expansion of the primitive vocabulary, which Carnap justifies by suggesting that “natural” is a logical term. Few have found this latter suggestion convincing, however.

Still, it is clear that criticisms of this sort threaten only *Aufbau*-style projects that involve phenomenal and/or logical bases. To avoid the problems, one need only expand the primitive basis. One can avoid Newman’s problem by allowing almost any nonlogical vocabulary. One can avoid Goodman’s problem by allowing expressions for specific sensory qualities. One can avoid Quine’s and Chisholm’s problems by allowing spatiotemporal expressions into the basic vocabulary directly, or perhaps by allowing expressions for causal relations.⁹

⁷See Newman 1928, and sections 153-155 of the *Aufbau*. Carnap marks these sections “can be omitted”, quite remarkably given the centrality of these sections to the “logical structure” project. For further discussion of Newman’s problem and the *Aufbau*, see Demopolous and Friedman 1985.

⁸In particular, as long as there is a property corresponding to any set of objects, and a relation corresponding to any set of ordered pairs, then the world-sentence S will be satisfied by any set of the right size. To see this, suppose that one set A of size n satisfies S , and let A' be any other set with the same size. Take a group of properties and relations that relate the members of A in the pattern specified by S . Map those properties and relations to a corresponding set of properties and relations on A' by a one-to-one mapping. (Any one-to-one mapping will do; the liberalness claim will ensure that every property maps to a property, and so on). Then the resulting properties and relations will relate the members of A' in the same pattern. So S will be satisfied by A . It follows that S cannot entail any truths that specify features of the world beyond its cardinality.

⁹Even while retaining a phenomenal base, Carnap has some options in avoiding the first three problems. As Leitgeb (2011) notes, Quine’s problem might be dealt with by invoking a Ramsey-sentence definition of external-world expressions in terms of phenomenal expressions, and Goodman’s problem might be dealt with by giving a more complex structural characterization of experiences as a basis (see also the discussion of phenomenal structuralism in 8.7). The Ramsey-sentence move might also help with Chisholm’s problem, although other general doubts about phenomenalist

These criticisms leave open, then, the possibility of pursuing an *Aufbau*-like project with a somewhat expanded basis. Carnap himself says that the choice of a phenomenal basis in the *Aufbau* is somewhat arbitrary, and that he could equally have started with a physical basis.¹⁰ To avoid the criticisms above, it suffices to invoke a basis with expressions for various sensory qualities and spatiotemporal relations. It is not clear that such a basis will fulfil all of Carnap's purposes in the *Aufbau*: it would not vindicate such a strong structuralism, and the epistemological consequences would be weaker. But many of the more general aims of a project of definability would be fulfilled.

Other doubts about the project of the *Aufbau* are more general. One influential source is Quine's critique of the analytic/synthetic distinction, also in "Two Dogmas of Empiricism". This critique has led many to doubt that a substantial distinction between the analytic and the synthetic, or between the a priori and the a posteriori, or between the definitional and the nondefinitional, can be drawn. If these doubts are correct, then any *Aufbau*-like project that involves these notions must fail.

Another source of doubt arises from Saul Kripke's arguments in *Naming and Necessity* against descriptivism: the thesis that names are equivalent to descriptions. Kripke argued that many expressions in natural language, including most names and natural kind terms, are neither a priori equivalent nor necessarily equivalent to descriptions using different expressions. Kripke's modal argument makes a case that for any name (e.g. 'Aristotle') and a typical associated description (e.g. 'the teacher of Alexander'), the name and the description are not necessarily equivalent. Kripke's epistemic argument makes a case that for any ordinary name (e.g. 'Gödel') and a typical associated description (e.g. 'the man who proved the incompleteness of arithmetic'), the name and the description are not a priori equivalent. If these claims are correct, then it appears that no *Aufbau*-like definitional project that applies to these terms and that invokes necessity or apriority constructions will remain.

¹⁰For many years, the popular conception of logical empiricism has focused on a commitment to phenomenalism and verificationism, and the *Aufbau* has been regarded as a paradigm of that tradition. In fact these theses do not play a central role in the *Aufbau*. A much more important role is played by Carnap's commitment to structuralism and objectivity, as discussed in chapter 8. The distortions in the popular conception are partly explained by a brief post-*Aufbau* period in the Vienna Circle in which phenomenal reductions involving protocol sentences played a more crucial role. Within a few years (e.g. Carnap 1932), however, Carnap had moved on to a view on which physical language rather than phenomenal language plays the crucial role in reduction. In recent years, the flourishing scholarly literature on the *Aufbau* (e.g. Friedman 1999 and Richardson 1998) has painted a picture that is much more nuanced than the caricature.

as a condition of adequacy can succeed.

Perhaps the most important source of doubt arises from general reasons to doubt that most expressions have definitions. The central problem is the problem of *counterexamples*. For many terms in English, it seems that every definition that has ever been offered is subject to counterexamples that render the definition inadequate.

The most famous case is the case of ‘knowledge’, traditionally defined as ‘justified true belief’. In 1963, Edmund Gettier pointed out that there are counterexamples to this purported definition. Suppose that Smith has a justified belief that Jones owns a Ford, and deduces that Jones owns a Ford or Brown is in Barcelona. And let us say that Jones has recently sold his Ford, and that Brown is in fact in Barcelona, though Smith has no information about either of these things. Then Smith has a justified true belief that Jones owns a Ford or Brown is in Barcelona, but this justified true belief is not knowledge. So knowledge cannot be defined as justified true belief.

In Gettier’s wake, others attempted to modify the definition to avoid these counterexamples, for example suggesting that knowledge can be defined as justified true belief that is not essentially grounded in a falsehood. But other counterexamples ensued: if I see the one real barn in an area of fake barns, and form the belief that I am seeing a barn, then this is a justified true belief not essentially grounded in a falsehood, but it is not knowledge. A parade of further attempted definitions and further counterexamples followed (see Shope 1983 for an exhaustive summary). Eventually definitions with fourteen separate clauses were proffered, with no end to the counterexamples in sight.

What goes for ‘knowledge’ seems to go for most expressions in the English language. Given any purported definition of ‘chair’, or ‘run’, or ‘happy’, it is easy to find counterexamples. For some scientific terms such as ‘gold’ or ‘electron’, there may be true definition statements (‘Gold is the element with atomic number 79’), but these do not appear to be a priori. For Wierzbicka’s definition of ‘lie’, above, counterexamples are not hard to find: I can tell a lie even if I do not care whether the hearer believes me.¹¹ And even in the case of ‘bachelor’, there are unmarried men who do not seem to be bachelors, such as those in long-term domestic partnerships. The only clearly definable expressions appear to be derived expressions (such as ‘unhappy’ and ‘caught’), which can arguably be defined in terms of the expressions (‘happy’ and ‘catch’) that they are derived from, along with some technical expressions that have been introduced through definitions, and a handful of others.

¹¹A philosopher will find possible counterexamples to many or most of Wierzbicka’s definitions. Of course it is likely that Wierzbicka’s intended criteria of adequacy for definitions differ from philosophers’ criteria, so it is not obvious to

The philosophical flight from definitions has been paralleled by a similar flight in cognitive science. Contemporary psychologists almost universally reject the so-called classical view of concepts, according to which most concepts are associated with sets of necessary and sufficient conditions. A major influence here is work by Eleanor Rosch and others on concepts such as that of a bird, suggesting that subjects classify various creatures as birds in a graded way according to their similarity to various prototypes rather than by necessary and sufficient conditions. In addition, psychologists often approvingly cite Ludwig Wittgenstein's discussion of the term 'game', suggesting that the concept of a game involves something more of a family resemblance between the things it applies to than a set of necessary and sufficient conditions. By and large, the classical view has been supplanted by views on which concepts involve prototypes, exemplars, and theories, among other views. On few of these views is it required that concepts are associated with definitions.

It remains possible that for these expressions, there exists an adequate definition that has not yet been found. In philosophy, the search for definitions typically runs out of steam once purported definitions reach a certain length. In psychology, it is not out of the question that prototype theories and the like might be used to deliver something like a definition, perhaps cast in terms of weighted similarities to certain prototypes or exemplars. Likewise, theory-based accounts of concepts might yield definitions of various concepts in terms of clusters of associated theoretical roles. Still, it is far from obvious that such definitions will exist, and even if they do exist, they will be so unwieldy that they will be quite unlike definitions as traditionally conceived. As a result, the definitional program has been put to one side in most areas of philosophy and psychology in recent years.

3 From Definitional Scrutability to A Priori Scrutability¹²

Even if Definitional Scrutability is false, there remains a strong case for other scrutability theses. For example, even if expressions such as 'knowledge' and 'chair' are not definable in terms of more primitive expressions, it remains plausible that there is some strong epistemological relation between truths involving these expressions and truths involving more primitive expressions. In particular, it is striking that in many cases, specifying a situation in terms of expressions that do not include 'knowledge' or its cognates (synonyms or near-synonyms) enables us to determine whether or not the case involves knowledge. Likewise, correctly describing an object in terms of

what extent the existence of counterexamples is a problem for Wierzbicka's project.

¹²This section overlaps in part with Chalmers and Jackson 2001.

expressions that do not include 'chair' or its cognates may enable us to determine whether or not it is a chair. And so on.

For example, in the Gettier situation we are told something like:

'Smith believes with justification that Jones owns a Ford. Smith also believes that Jones owns a Ford or Brown is in Barcelona, where this belief is based solely on a valid inference from his belief that Jones owns a Ford. Jones does not own a Ford, but as it happens, Brown is in Barcelona.'

Let the conjunction of these sentences be G . G does not contain the term 'know' or any cognates. But when presented with G , we are then in a position to determine that the following sentence K is false:

'Smith knows that Jones owns a Ford or Brown is in Barcelona'.

Something like this happens throughout philosophy, psychology, and other areas. We are given a description D of a scenario without using a key term E , and we are asked to determine whether and how the expression E applies to it. This is the key method for experimental work on concepts in psychology: an experimenter presents a description (or perhaps a picture) of a case, subjects are asked to classify it under a concept, and they very frequently can do so. The same goes for conceptual analysis in philosophy: one considers a specific case, considers the question of whether it is a case of an F , and one comes to a judgment. Often we have no trouble doing so.

In fact, this is precisely how counterexamples to definitions are often generated. When someone suggests that E can be defined as F (say, bachelors are defined as unmarried men), someone else suggests a scenario D (say, involving long-term gay couples) to which F applies but E does not, or vice versa. The Gettier case fits this pattern perfectly. Despite the absence of definitions, there is some form of scrutability present in these cases: once we know G , we are in a position to know $\neg K$, and so on. And in fact, the argument from counterexample relies on this sort of scrutability.

In many cases, it is plausible that the scrutability is a priori. For example, in the Gettier case, it is plausible that one can know the material conditional 'If G , then $\neg K$ ' a priori. Someone who knows that G is true and who possesses the concepts involved in K (in particular the concept of knowledge) is thereby in a position to know that K is false, even if they lack any further relevant empirical information. That is, a grasp of the concept of knowledge (along with a grasp of the

other concepts involved) and rational reflection suffices to eliminate the possibility that both G and K are true.

On the face of things, Gettier's argument was an a priori argument, in which empirical information played no essential role, and its conclusion is a paradigmatic example of a non-obvious a priori truth. The argument proceeds by presenting the hypothesis that G holds, and appealing to the reader's possession of the concept of knowledge to make the case that if G holds, $\neg K$ holds (and J holds, where J is a corresponding positive claim about Smith's justified true belief). Empirical information plays no essential role in justifying belief in this conditional, so the conditional is a priori. The a priori conditional itself plays an essential role in deriving the a priori conclusion.

This brings out an important point: a priori scrutability does not require definability. One might think that for a sentence B to be a priori entailed by a sentence A , the terms in B must be definable using the terms of A . However, this thesis is false. The a priori entailment from 'There exists a red ball' to 'There exists a colored ball' is one counterexample: 'colored' cannot be defined in terms of 'red'. But the case above is another counterexample. At least once general skepticism about the a priori is set aside, 'If G then $\neg K$ ' is a typical example of an a priori truth. But at the same time, we have seen that there is little reason to think that there is an adequate definition of 'knowledge', whether in the terms involved in G or any other terms.

As before, it could be that there is an adequate definition that has not yet been produced, or that has been produced but overlooked. Someone might even hold that all these a priori conditionals are underwritten by our tacit grasp of such a definition. But even if so, it seems clear that the a priori entailment from G to K is not dialectically hostage to an explicit analysis of knowledge that would support the entailment. That is, we can have reason to accept that there is an a priori entailment here even without having reason to accept that there is an explicit analysis that supports the entailment.

If anything, the moral of the Gettier discussion is the reverse: at least dialectically, the success of a definition itself depends on a priori judgments concerning specific cases, or equivalently, on a priori judgments about certain conditionals. The Gettier literature shows repeatedly that purported definitions are hostage to specific counterexamples, where these counterexamples involve a priori judgments about hypothetical cases. So a priori conditionals seem to be prior to definitions at least in matters of explicit justification. Our judgments about a priori conditionals do not need judgments about definitions to justify them, and are not undermined by the absence of definitions.

It might be suggested that our conditional judgments here require at least explicit *sufficient* conditions for knowledge or its absence: for example, the condition that a belief based solely

on inference from a false belief is not knowledge. Of course it is trivial that there is a sufficient condition in the vicinity of such an entailment (the antecedent provides one such), so the claim will be interesting only if the complete set of sufficient conditions for knowledge is not huge and open-ended. But the Gettier literature suggests precisely that the set of sufficient conditions for knowledge is open-ended in this way; if it were not, we would have a satisfactory definition. And as before, the a priori entailments are not dialectically hostage to the proposed sufficient conditions; if anything, proposed sufficient conditions are hostage to a priori intuitions about specific cases.

It is far from clear that in every case of knowledge, there will be some short sufficient condition for knowledge that obtains in that case. It may be that any reasonably short condition not involving ‘knowledge’ or cognates is compatible with the absence of knowledge.¹³ This is clearly not so for every predicate: there are plausibly short sufficient conditions for *not* knowing that *p*, such as not believing that *p*, or believing that *p* based solely on inference from a false belief. But it may be that for many predicates, there are at least hypothetical cases for which there is no reasonably short sufficient condition (perhaps even no finite sufficient condition) obtaining in that case. These observations suggest that for some sentences and some scenarios, a priori scrutability of these sentences may require very many or very long sentences in the base. Of course this is compatible with the scrutability thesis, but it reinforces the need for idealization in understanding the thesis.

It remains plausible that the expressions in question have *approximate definitions*. There is an intuitive sense in which ‘justified true belief’ is a fairly good approximate definition for ‘knowledge’: it gets most cases right, in an intuitive sense of ‘most’. ‘Justified true belief not essentially grounded in a falsehood’ is even better. In the face of counterexamples, one can refine definitions yielding longer and longer definitions that cover more and more cases. But it is not obvious that there will be any finite definition that gets all possible cases right (see 8.2 for more on this). There

¹³See Williamson 2000 for discussion of this point in the context of knowledge. Williamson 2007 suggests that common descriptions of Gettier cases do not suffice for the absence of knowledge, for example because there are cases compatible with these descriptions in which subjects have other evidence for the relevant *p* (see Malmgren 200x and Ichikawa and Jarvis 200x for discussion). I think that *G* above might escape this charge, because it includes the “based solely on” clause. But the point will still apply to justification: there will be deviant cases that satisfy *G* but not *J*, because extraneous factors undermine Smith’s justification for believing the relevant proposition. One might still argue that an inference from *G* to *J* or a conditional belief in *J* given *G* is justified a priori (in the nonconclusive sense of apriority discussed in 4.1), but in light of the deviant cases it is not straightforward to maintain that a belief in the material conditional from *G* to *J* is justified a priori. Still, if one fills out *G* to include all relevant detail—in the limit moving to the sort of complete scenario specification discussed in 3.1—worries of this sort will be removed.

may be some infinite definition (one could presumably enumerate all possible cases of knowledge in an infinite disjunction, for example), but such a definition would almost certainly fall short of the criteria of adequacy typically imposed on a good definition. Likewise, even if there is some very long finite definition that gets all the clear cases right (perhaps setting aside the desideratum that a good definition should classify unclear cases as unclear), it seems unlikely that this will meet other standard criteria of adequacy for definitions. But as before, accepting a priori scrutability does not require accepting that any such definition exists.

A useful model here holds that even if expressions do not have definitions, they still have *intensions*: functions that map possible scenarios to extensions. So ‘*S* knows that *P*’ has an intension that classifies possible scenarios according to whether the sentence is true or false there, ‘bachelor’ has an intension that maps possible scenarios to the class of bachelors in that scenario, and so on. For our purposes, we can think of these scenarios as *epistemically possible* scenarios: roughly, highly specific ways the world might turn out that we cannot rule out a priori. (Note that here and throughout, epistemic possibility is tied to what can be known a priori rather than to what a given subject knows.) For a given scenario, we can consider the hypothesis that the scenario actually obtains, and come to a judgment about the extension of an expression (such as ‘*S* knows that *P*’ or ‘bachelor’) with respect to that scenario. For a few expressions, we might be able to capture the intension of the expression in the form of a definition. For many expressions, however, we can only approximate the intension with a definition, giving better and better approximations with longer and longer approximate definitions.

On this model, one might suppose that judgments about cases are underwritten by a sort of grasp of the intension of an expression, rather than by a grasp of the definition of an expression. This grasp of an intension might itself correspond to a sort of *conditional ability* to identify an expression’s extension, given sufficient information about how the world turns out and sufficient reasoning. That is, a sufficiently rational subject using expressions such as ‘bachelor’, ‘knowledge’, and ‘water’ will have the ability to evaluate certain conditionals of the form $E \rightarrow C$, where *E* contains relevant information about the world (typically not involving the expression in question) and where *C* is a statement using the expression and characterizing its extension. And in order that it is not an accident that subjects can do this in the actual world, subjects will also be able to do this given specifications of many different scenarios.

Arguments from counterexample can make a case against definitions, but they cannot make a case against the claim that expressions have intensions. Such arguments themselves proceed by considering scenarios (say, a Gettier scenario), and by making the case that the extension of

an expression E (' S knows that P ') with respect to that scenario differs from the extension of a purported definition D (' S has a justified true belief that P '). To capture the intuitive data here, we need only suppose that the intension of the expression picks out the intuitive extension at that scenario (in this case, false) rather than the intuitive extension of the definition (in this case, true).

To see this in more detail, first note that arguments from counterexample can be used both to oppose the claim that $E == D$ is necessary, and to oppose the claim that $E == D$ is a priori. The first sort of argument requires exhibiting a *metaphysically possible* situation (roughly, a situation that might have obtained) of which the equivalence is false. For example, it seems that the Gettier situation is metaphysically possible, in that it could have obtained. Furthermore, if it had obtained, then there would have been justified true belief without knowledge. It follows that it is not necessary that knowledge is justified true belief. The second sort of argument requires exhibiting an *epistemically possible* scenario (that is, a scenario not ruled out a priori) of which the equivalence is false. For example, it seems that the Gettier situation is epistemically possible, in that it is not ruled out a priori that it obtains. Furthermore, if it obtains, then there is justified true belief without knowledge. It follows that it is not a priori that knowledge is justified true belief.

Kripke's modal argument against descriptivism is an argument from counterexample of the first kind. It focuses on a metaphysically possible situation in which Aristotle did not go into pedagogy, and makes the case that if this situation had obtained, then Aristotle would not have been the teacher of Alexander. It follows that it is not necessary that Aristotle was the teacher of Alexander. Kripke's epistemological argument against descriptivism is an argument from counterexample of the second kind. It focuses on an epistemically possible situation in which the proof of the incompleteness of arithmetic was stolen, and makes the case that if that situation actually obtains, then Gödel is not the prover of incompleteness. It follows that it is not a priori that Gödel is the prover of incompleteness.

In effect, the first sort of argument shows that the *modal profile* of an expression (the way it applies across metaphysically possible worlds) is not identical to that of a purported definition. Such an argument is clearly compatible with the claim that the modal profile can be represented as an intension, however. Likewise, the second sort of argument shows that the *epistemic profile* of an expression (the way that it applies across epistemically possible scenarios) is not identical to that of a purported definition. Again, such an argument is clearly compatible with the claim that the epistemic profile of an expression can be represented as an intension.¹⁴

¹⁴In the case of an expression such as 'knowledge', the modal and epistemic profiles appear to be more or less the same, so one intension will suffice to represent both. In the case of names such as 'Aristotle' and 'Gödel', the modal

It is clear that modal arguments by counterexample do not make any case against the thesis that ‘knowledge’-truths and the like are *necessitated* by a description of a situation in more basic terms. Likewise, epistemic arguments by counterexample do not make any case against the thesis that ‘knowledge’-truths and the like are *a priori entailed* by a description of a situation in more basic terms. That is, no such argument can make a case against A Priori Scrutability.

In fact, there is a reasonable *prima facie* case for A Priori Scrutability here. Certainly there will be many conditionals $E \rightarrow C$, with relevant information in the antecedent, that are a posteriori. For example, it is a posteriori that if a glass contains H₂O, it contains water. But these will be cases in which the antecedent E does not contain sufficient empirical information to identify the expression’s extension given possession of the concept and reasoning alone. To justify belief in the conditional, further empirical information F is needed: for example, information regarding the appearance, behavior, and distribution of H₂O molecules in the world. But then we can simply conjoin E with F , yielding E' , and we will obtain a new conditional $E' \rightarrow C$. It is plausible that because F is built into the antecedent of this conditional, F will not be required to justify the conditional. If so, and if F specifies all relevant empirical information, then the conditional will be a priori.¹⁵

In these cases, all truths involving the key expression will be scrutable from truths not involving that expression. In effect, the expression can be eliminated from an a priori scrutability base. Of course this elimination cannot be extended forever. There may be some expressions (‘and’, ‘conscious’, ‘exists’?) that are so basic that in order for $E \rightarrow C$ to be a priori, where C involves these expressions and E specifies relevant information, E must itself include these expressions or close cognates. Still, it is plausible that if the model above works for ‘knowledge’ and ‘water’, it will work for enough expressions that the antecedents of these a priori conditionals can be restricted to a quite limited class of expressions.

If the antecedents can be restricted to a compact class of expressions, then we will have vindicated the A Priori Scrutability thesis, as formulated earlier:

A Priori Scrutability: There is a compact class of truths such that all truths are a priori scrutable from truths in that class.

profiles and the epistemic profiles may be quite distinct, so one needs distinct intensions to represent them. These are just the primary and secondary intensions of two-dimensional semantics (discussed in 5.5 and E10). For present purposes, it is the epistemic profile (and the corresponding intension) that is most important.

¹⁵This sort of argument is developed at greater length in 4.4.

From now on, my focus will largely be on A Priori Scrutability rather than Definitional Scrutability. When I speak of “the scrutability thesis”, it is A Priori Scrutability that should be understood, and when I say that one class of truths is scrutable from another, it is the a priori scrutability relation that should be understood.

I do not claim to have established A Priori Scrutability here. I have just given some motivation for taking it seriously. I will argue for the thesis at much greater length in chapters 3 and 4, after clarifying the thesis and related scrutability theses in chapter 2. Still, I hope to have said enough to make clear that the falsity of Definitional Scrutability is compatible with the truth of A Priori Scrutability, and that the arguments by counterexample that pose serious objections to the Definitional Scrutability thesis have no corresponding force against the A Priori Scrutability thesis. If anything, reflection on what is going on in these cases tends to support the A Priori Scrutability thesis.

It is also worth noting that Kripke’s arguments against descriptivism, which pose significant problems to Definitional Scrutability when applied to names and natural kind terms, do not have similar force when applied to A Priori Scrutability. This is especially clear in the case of Kripke’s modal argument, which concerns what is metaphysically possible and necessary, matters about which the scrutability thesis makes no claims. The scrutability thesis merely makes claims about what is a priori and a posteriori. Kripke’s modal argument allows that a conditional such as ‘If X is the dominant cause of heat sensations, X is heat’ can be a priori even when it is not metaphysically necessary. So there is no objection to the scrutability thesis here.

Kripke’s epistemological argument targets the claim that ordinary names are a priori equivalent to descriptions. But just as A Priori Scrutability is entirely compatible with the claim that ordinary expressions have no explicit definition, it is compatible with the claim that ordinary names are not a priori equivalent to descriptions. And we have seen that arguments from counterexample such as this cannot make any case against scrutability. For example, Kripke’s argument concerning ‘Gödel’ gives no reason to deny that sentences such as ‘Gödel did not prove incompleteness’ are themselves scrutable from a specification of the relevant scenario. In fact, one might see the arguments themselves as providing yet another example of scrutability reasoning.

This relationship between scrutability and Kripke’s arguments has significant consequences. Kripke’s arguments are often thought to undermine broadly Fregean analyses of meaning and content. But we will see shortly, and in much more detail later in this book, that the scrutability thesis can be used to support a broadly Fregean analysis of meaning and content. So if I am right that Kripke’s arguments do not undermine the scrutability thesis, these arguments are more limited

in scope than is often supposed.

4 The Scrutability Base

A *scrutability base* is a class of truths from which all truths are scrutable, for a given notion of scrutability. What sort of truths might go into a scrutability base?

At the end of the *Aufbau*, Carnap embraced what we might call Logical Scrutability: the view that there is a scrutability base using only logical expressions. Some phenomenologists accept Phenomenal Scrutability, holding that there is a scrutability base using only phenomenal expressions (expressions characterizing the character of conscious experiences) and logical expressions. Some physicalists accept Microphysical Scrutability, holding that there is a scrutability base using only microphysical expressions (expressions used in fundamental physics) and logical expressions.¹⁶

For our purposes, all of these views are strong and interesting scrutability theses (versions of all of them are entertained by Carnap in the *Aufbau*), but the current project is not committed to any of them. Our central scrutability thesis is what we might call Compact Scrutability (often just “the scrutability thesis” for short): there is a compact class of truths such that all truths are scrutable from truths in that class. Given that logical, microphysical, and phenomenal bases count as compact, then Logical, Phenomenal, and Microphysical Scrutability entail Compact Scrutability. But less austere bases than these may still be compact.

What is compactness, exactly? As I characterized compactness earlier, a class of truths is compact iff it uses only a small class of expressions. A little more precisely, we can say that compactness requires that a class of truths uses only expressions from a small number of *families* of expressions. If it turns out that all truths are scrutable from phenomenal truths, but that an infinite number of phenomenal expressions are required to capture the diversity of possible phenomenal qualities, this would still be a strong enough scrutability thesis for our purposes. We can stipulate that the class of phenomenal expressions counts as a single family, as does the class of microphysical expressions, the class of logical expressions, the class of mathematical expressions, and so on. The intuitive idea here is that expressions in the same family should share a common domain. (So

¹⁶In principle, these views concerning a scrutability base can be combined with different notions of scrutability, yielding such theses as Definitional Phenomenal Scrutability, A Priori Physical Scrutability, and so on. Carnap’s versions of these theses are the definitional versions. For the most part my focus will be on a priori scrutability and on compact bases, so the A Priori and Compact versions should be understood where not otherwise indicated. For more on the conventions here, see 2.1.

the class of spatiotemporal expressions counts as a family, while the class of singular terms does not.) Beyond this I will leave the notion of a family intuitive.

How small is small? We can leave this notion vague. But to give a rough idea, I would say that fewer than ten or so families would be ideal, that twenty would be acceptable, but that more than a hundred would be pushing things. One could also stipulate that a compact class of truths will exclude the great majority of terms used in natural languages: there will be few or no ordinary proper names ('London', 'George Bush'), natural kind terms ('water', 'kangaroo'), artifact terms ('car', 'table'), and neither will there be cognate terms in a different language, constructions from such terms, and so on. The idea is that truths involving terms like this should all be scrutable from truths in a more primitive vocabulary. I will not build this into the official definition, but one can see this as part of the spirit of the thesis.

It is worth noting that while a compact class of truths must use only a limited vocabulary, it need not include *all* truths that use a given vocabulary. For example, there is a compact class of truths that includes all microphysical truths but not all mathematical truths. Stating the microphysical truths may require mathematical vocabulary, but many truths that use only mathematical vocabulary will not be included.

We also need to require that a compact class of truths avoids *trivializing mechanisms*. There are certain sorts of base truths that threaten to render the scrutability thesis trivial. One such is a base consisting of the family of expressions for *propositions*, along with 'is true'. It is not implausible that every sentence is scrutable from a sentence saying that a corresponding proposition is true, but this result is not interesting. Likewise, one could perhaps code all truths of English into a single real number ϕ , via an appropriate coding scheme: then it is not implausible that all such truths are scrutable from the single truth that ϕ equals such-and-such. But again, this thesis is not interesting. There is a clear sense in which these proposals involve trivializing mechanisms, by somehow directly coding a large number of truths from different families into a single truth or a single family of truths. I will not attempt to define this notion, but it should be understood that compact classes cannot include sentences of this sort.

So a class of sentences is compact if it includes expressions from only a small number of families and includes no trivializing mechanisms. Of course this notion is vague and has not been precisely defined. But in practice, this will not matter. The sort of specific scrutability claims I will discuss and defend will all involve highly restricted vocabularies that are clearly small enough to be interesting. In most cases, there will be no threat of a trivializing mechanism, and when there is such a threat, it can be discussed directly.

How small can a scrutability base be? Let us say that a *minimal* scrutability base is a class of sentences C such that (i) C is a scrutability base, (ii) no proper subclass of C is a scrutability base, and (iii) there is no scrutability base using only a proper subclass of the expressions used in C .

Three proposals about minimal scrutability bases are given by the theses of Logical Scrutability, Phenomenal Scrutability, and Microphysical Scrutability. I think that there are good reasons to reject these proposals, however. In part for reasons we have already discussed, it is plausible that many physical truths are not a priori scrutable from logical or phenomenal truths. Conversely, it is plausible that many phenomenal truths are not a priori scrutable from a microphysical base. For example, it appears that no amount of a priori reasoning from microphysical truths will settle what it is like to see red (Jackson 1982). This suggests that many phenomenal truths (truths concerning the character of conscious experiences) are not a priori scrutable from microphysical truths. It also appears that no amount of a priori reasoning from microphysical truths will enable one to know such perspectival truths as ‘It is now March’, or such negative truths as ‘There are no ghosts’.

Still, this leaves more liberal scrutability theses on the table. For example, Chalmers and Jackson (2001) suggest that all ordinary macroscopic truths are a priori entailed by the class of physical *and* phenomenal truths along with certain indexical truths (‘I am such-and such’, ‘Now is such-and-such’) and a “that’s-all” truth (on which more in 3.1 and E5). If this class of truths a priori entails *all* truths (as I argue in chapters 3, 4, and 6), then it can serve as a scrutability base. There may be even smaller bases. For example, microphysical truths may themselves be scrutable from a base involving phenomenal expressions and nomic expressions (such as ‘law’ or ‘cause’), perhaps along with spatiotemporal and/or mathematical expressions. If so, then (as I argue in chapter 7) a scrutability base might need to involve only phenomenal, nomic, logical, and indexical expressions,¹⁷ perhaps along with spatiotemporal and/or mathematical expressions. On some views (explored in chapters 7 and 8), the base may be smaller still.

A few principled scrutability bases are worthy of attention. One base, in the spirit of Carnap’s own view, yields the thesis of Structural Scrutability: all truths are scrutable from structural truths. If structural truths are restricted to a logical vocabulary, this view falls prey to Newman’s problem. But we might understand structural truths more expansively, to let in truths about fundamentality or naturalness (as on Carnap’s own final view), or about laws and causation, for example. I explore the viability of views of this sort in Chapter 8.

¹⁷Throughout this book, I count as “indexical expressions” just a limited class of perspectival expressions: ‘I’, ‘now’, and perhaps certain heavily constrained demonstratives. In this sense, indexical expressions count reasonably as a family. I use “context-dependent” for the broader class of expressions whose content depends on context.

Another principled scrutability thesis is Fundamental Scrutability: the thesis that all truths are scrutable from metaphysically fundamental truths (plus indexical truths and a that's-all truth, if necessary). The metaphysically fundamental truths are those that serve as the metaphysical grounds for all truths: they might involve attributions of fundamental properties to fundamental entities.¹⁸ On a standard physicalist view, the metaphysically fundamental truths are microphysical truths. On a standard property dualist view, metaphysically fundamental truths may include microphysical and phenomenal truths.¹⁹

Another thesis, in the spirit of Russell's quite different constructions of the world, is Acquaintance Scrutability: all truths are scrutable from truths about entities with which we are directly acquainted. Another, in the spirit of the thesis about concepts with which we started this chapter, is Primitive Scrutability: all truths are scrutable from truths involving only expressions for primitive concepts. Yet another, relevant to debates about internalism and externalism about meaning and content, is Narrow Scrutability: all truths are scrutable from truths whose content is determined by the internal state of the subject. In chapter 8, I will make a case for all three of these theses, as well as a tentative case for Fundamental Scrutability.

Some potential scrutability bases are less austere than others. For example, someone might think that we need normative expressions ('ought') in the base, or that we need expressions for secondary qualities ('red') in the base, or that we need intentional notions ('believes') in the base. If a scrutability base needs to be expanded to include these expressions, then the base will plausibly go beyond the structural or the metaphysically fundamental, but it will still be small enough that we will have a strong and interesting scrutability thesis.

Of course there is more than one scrutability base. For a start, as long as scrutability is monotonic (if S is scrutable from C , S is scrutable from any set of truths containing C) adding truths to any scrutability base will yield a scrutability base, and substituting a priori equivalent synonyms within a scrutability base will also yield a scrutability base. Even if we restrict ourselves to minimal scrutability bases (scrutability bases of which no proper subclass is a scrutability base, and such that there is no scrutability base using a proper subset of the expressions) and factor out synonyms, a diversity of bases is possible. For example, given a minimal scrutability base involving predicates F and G , there will also be a minimal scrutability base involving four new predicates

¹⁹Metaphysical fundamentality should be distinguished from conceptual primitiveness. One might reasonably hold that spin and charge are metaphysically fundamental without holding that the concepts *spin* and *charge* are primitive. Likewise, one might hold that the concept *I* is primitive without holding that the self is anything metaphysically fundamental. Still, there may be an attenuated relation between the two; see E16, and also 8.4 and 8.6.

H, I, J, and K, corresponding to conjunctions of *F, G*, and their negations. One can also obtain multiple bases from the familiar idea that there can be a priori equivalent formulations of a physical theory in different vocabularies. It is even not out of the question that on some views, both a microphysical vocabulary and a phenomenal vocabulary (or a phenomenal vocabulary combined with a nomic or spatiotemporal vocabulary) could yield minimal scrutability bases.

For most of our purposes, the existence of multiple scrutability bases is not a problem. Still, the phenomenon does suggest that the mere fact that an expression is involved in a minimal scrutability base does not suffice for that expression to be a primitive concept in any interesting sense. And there remains an intuition that some scrutability bases are more fundamental than others. For example, in the case above, it is natural to hold that predicates *F* and *G* stand in certain conceptual, epistemological, and psychological priority relations to *H, I, J, and K*. Likewise, one might hold that phenomenal and nomic expressions stand in certain conceptual, epistemological, and psychological priority relations to microphysical expressions. This will be especially clear if one holds that microphysical expressions are definable in terms of phenomenal and nomic expressions, but even if one rejects the definitional claim, one might still accept some sort of priority claims.

I take the moral here to be that a priori scrutability is a relatively coarse-grained relation among classes of truths. One might react to this moral by postulating a more fine-grained relation of dependence among truths (perhaps some sort of conceptual dependence), such that whenever one class of truths depends on another, truths involving the former are scrutable from truths involving the latter, but not vice versa. On this way of doing things, many scrutability bases will not be dependence bases, and it is not out of the question that there might be just one minimal dependence base (at least up to equivalence through synonymy).²⁰

This line of thought immediately raises the question of how the fine-grained dependence relation in question should be understood. If one accepts the definitional model, one might suggest that the relation is just definitional scrutability, and that the dependence base will involve all and only the undefinable expressions. But if one rejects the definitional model, the correct understanding is less clear. Later in this book (chapter 9 and E16), I explore another way of understanding such a fine-grained relation, along with the relationship between fine-grained relations and scrutability.

For now, I will concentrate on priori scrutability and related coarse-grained notions. These have the advantage of being better-understood than more fine-grained notions, so that arguing for

²⁰This reaction is an epistemological or conceptual analog of a familiar metaphysical line of thought concerning supervenience, leading some to postulate relations of ontological dependence or grounding that are finer-grained than the coarse-grained relation of supervenience. For more on these issues, see E16.

scrutability theses of this sort is more straightforward. A number of the scrutability bases I will consider will also be plausible candidates to be dependence bases, so that the expressions involved will be plausible candidates to be primitive concepts. But even in the absence of claims about dependence and primitiveness, these scrutability theses have significant consequences.

5 Scrutability and the *Aufbau*

If the A Priori Scrutability Thesis is correct, it offers a vindication of *something like* the project of the *Aufbau*.²¹ There are two significant differences: the very limited bases (logical and/or phenomenalist) of the *Aufbau* are replaced by somewhat less limited bases here, and the role of definitional entailment in the *Aufbau* is played by a priori entailment here. The expansion of the base allows us to avoid Goodman's, Quine's, and Chisholm's objections to the phenomenalist base, and Newman's objection to the purely logical base. The move from definitions to a priori entailment allows us to avoid the central problems for definitions and descriptions, including the problem of counterexample, and Kripke's modal and epistemological arguments against descriptivism.

Of course there are challenges to the *Aufbau* that also apply to the scrutability thesis. Most notably, Quine's critique of the analytic/synthetic distinction is often thought to generate an equally significant critique of the a priori/a posteriori distinction, and so has the potential to undermine the a priori scrutability thesis. In chapter 5, however, I will suggest that an analysis in terms of scrutability provides the materials required to show where Quine's arguments go wrong. I will address a number of other challenges to the scrutability thesis in chapters 3 and 4.

One might ask: does the scrutability thesis have the potential to satisfy some of the ambitions of the *Aufbau*? These ambitions included an analysis of meaning and concepts, a sort of epistemological optimism, a sort of metaphysical deflationism, and a language that might help to unify science. These elements were supposed to jointly yield a sort of blueprint for scientific analysis and philosophical progress. Of course the *Aufbau* is widely held to have failed in these ambitions, and I will not try to put anything so strong in their place. Still, the scrutability thesis has consequences in many different areas of philosophy, consequences that share at least some of the flavor

²¹A quite different project in a similar spirit, attempting to vindicate something like the *Aufbau*, is carried out by Hannes Leitgeb (2011). Leitgeb retains a phenomenal basis, although he gives it more structure than Carnap allowed. He also retains definitional entailment by imposing a relatively weak criterion of adequacy according to which definitions must involve "sameness of empirical content". On this criterion, definitions can be false. Because of this, I think that Leitgeb's version of the *Aufbau* will not play the semantic, metaphysical, and epistemological roles that I am interested in, but it may well be able to play other roles.

of Carnap's ambitions in the *Aufbau* and other works.²²

1. *Knowability and skepticism.* The notorious *Knowability Thesis* in epistemology, often associated with the programs of logical empiricism and verificationism, holds that all truths are knowable. This thesis is now widely rejected, for both formal and intuitive reasons. I argue shortly (E1) that scrutability theses can capture at least a plausible relative of these theses that can play some elements of the role that the knowability thesis has been used to play. Furthermore, certain scrutability theses offer a distinctive response to skepticism (E13).

2. *Modality.* Carnap's *Aufbau* project yields a basic vocabulary that can be used not just to characterize the actual world, but also other possible states of the world. This leads directly to Carnap's later project in *Meaning and Necessity* (1947), in which he analyses possibility and necessity in terms of state-descriptions for other possible worlds. While this sort of construction is now often used to understand metaphysically possible worlds, the scrutability framework allows such a construction to yield a space of *epistemically* possible worlds, or scenarios (E9). One can use a generalized scrutability thesis to define epistemically possible scenarios in terms of maximal a priori consistent sets of sentences in a scrutability base. These are analogous to Carnap's state-descriptions (possible worlds), and behave in a more Carnapian way than possible worlds on the usual contemporary understanding. For example, a posteriori sentences such as 'Hesperus is Phosphorus', are true in all metaphysically possible worlds, but they are false in some epistemically possible scenarios, as one might expect. So these scenarios can play a role in analyzing epistemic possibility analogous to the role of possible worlds in analyzing metaphysical possibility.

3. *Meaning.* Carnap's construction in *Meaning and Necessity* was intended to support a Fregean analysis of meaning, by understanding meanings as intensions defined over possible worlds. As discussed in chapter 5 and especially E10, the scrutability framework can be used to help vindicate this Fregean project by defining intensions over epistemically possible scenarios as above. For example, one can define the (epistemic or primary) intension of a sentence as the

²²For more on these applications, see E1 and E13 (knowability and skepticism, respectively), E9 and E10 (modality and meaning), 8.3 and 8.4 (primitive concepts and narrow content), 8.6 and E16 (metaphysics), 8.7 and E11 (structuralism and the unity of science), and 6.5 and 9.4 (metaphilosophy). It should be noted that many of these applications require specific scrutability theses. For example, the reply to skepticism requires Structural Scrutability or a variant thereof. The analysis of narrow content requires Narrow Scrutability. Central applications to metaphysics require theses such as Fundamental Scrutability. The crucial applications to meaning and modality require less, but they work better if one at least has scrutability from a compact base consisting of non-context-dependent expressions and primitive indexicals, and better still if one has a version of Acquaintance Scrutability. See chapter 8 for a discussion of most of these matters.

set of scenarios in which it is true. Then two sentences will have the same intension iff they are a priori equivalent. One can go on to define intensions for other expressions, such as singular terms, such that 'a' and 'b' will have the same intension iff 'a=b' is a priori. So 'Hesperus' and 'Phosphorus' will have different intensions. If the scrutability thesis is true, intensions of this sort will behave in a manner reminiscent of Fregean sense.

4. *Concepts and mental content.* In the *Aufbau*, Carnap put much emphasis on the construction of concepts. We can use the scrutability framework to associate intensions not just with linguistic items such as sentences but with mental items such as thoughts. As in the case of language, these intensions will serve as contents that reflect the epistemological properties of thoughts. Under some reasonable assumptions (outlined in the discussion of Narrow Scrutability in chapter 8), these intensions can also serve as *narrow* contents of thought: contents that are wholly determined by the intrinsic state of the thinker. These contents, grounded in a priori inferential relations to thoughts composed of primitive concepts, can go on to ground wide contents in turn. This approach to content naturally leads to a view in which primitive concepts play a grounding role with respect to all intentionality, and suggests that the path to naturalizing intentionality may proceed through the naturalization of the content of these primitive concepts.

5. *Metaphysics.* With specific scrutability theses in hand, these theses can be used to argue for realism or anti-realism about a given subject matter. For example, given Fundamental Scrutability, the thesis that all truths are scrutable from fundamental truths, then if ontological sentences (sentences about the existence of composite objects, say) are not scrutable from more fundamental truths, then they are either themselves fundamental or they are not true. The same goes for moral sentences, for sentences about consciousness, and so on. Furthermore, if we can use principled reasons and arguments to find a correct scrutability base, it will follow that any sentence that is not scrutable from this base is not true. Depending on just what is in the base, certain forms of anti-realism about the relevant domain may follow.²³

6. *Scientific analysis.* The unity of science was one of the major concerns of the logical empiricists, and Carnap hoped that the *Aufbau* program might contribute to this unity by showing how

²³This application is restricted to distinctions between realism and anti-realism that can be drawn in terms of truth and falsity. The framework does not bear so directly on distinctions that are drawn differently: for example, arguments of this sort will not easily distinguish moral realism from varieties of moral anti-realism that allow that 'Such-and-such is good' is true. The framework itself is largely neutral on the nature of truth and its grounds in various domains. While I lean toward a correspondence view of truth myself, the arguments of this book are compatible with many different analyses of both realist and anti-realist flavors.

all scientific notions could be analyzed in terms of a common basic vocabulary. If the scrutability thesis is true, then all scientific truths are at least scrutable from a common base. Furthermore, it can be argued that when scientific truths are scrutable from other truths of which there is a scientific account, then this account can be used to provide an explanation of the scrutable truths. If so, then (as I argue in E11), scrutability might yield a relatively unified account of all scientific truths. Scrutability also helps to analyze the prospects for structuralist views of science (8.7).

7. *Metaphilosophy*. The scrutability thesis entails that all philosophical truths are scrutable from base truths. So even philosophical ignorance can be localized to our ignorance of base truths or the non-ideality of our a priori reasoning (6.5). An extension of the scrutability thesis (chapter 9) suggests a way of reducing all philosophical disagreements to disagreements over base truths.

Of course the analysis of meaning and concepts that one gets from this project is more open-ended than in the ambitions of the *Aufbau*, the epistemological optimism is attenuated, and any metaphysical deflationism is more limited. Still, the consequences are strong and striking enough that the scrutability thesis is certainly worthy of investigation.

First Excursus: Scrutability and Knowability

The scrutability thesis is related to a number of widely-discussed theses in analytic philosophy. In this excursus, I discuss the relation to the knowability thesis and its cousin the verification principle. In the next excursus, I discuss its relation to Quine's thesis of the inscrutability of reference. Doing so can help to indirectly motivate the scrutability thesis, by showing how it avoids problems for related theses while still capturing something of their flavor.

First, the Knowability Thesis.

Knowability Thesis: For any truth S , it is possible that someone knows S .

This thesis is often doubted, for both intuitive and formal reasons. Intuitively, it seems that there may be truths concerning the distant past, the far away, and the very small, that it may be impossible for anyone to know. Formally, the thesis gives rise to what is often known as the Paradox of Knowability, first published by Fitch (1963).

Fitch in effect gives a disproof of the Knowability Thesis, arguing from the weak assumption that some truth is not *known* to the conclusion that some truth is not knowable. Let P be a truth such that in the actual course of history, no-one ever knows P . Let Q be ' P and no-one knows that P '. Then Q is true, but Q is unknowable. If someone were to know Q , then they would know P , but if someone were to know P , then Q would be false. So no-one can know Q .

The scrutability thesis is closely related to the knowability thesis. It does not say that every truth is knowable, but it does say that every truth is scrutible, or derivable from a limited class of basic truths. One might thereby wonder whether scrutability theses are liable to similar problems. In the introduction, we saw briefly that Inferential Scrutability is liable to problems related to Fitch's paradox, problems that I discuss at more length in 2.4 and 3.5. However, Conditional and A Priori Scrutability avoid both sorts of problems.

Concerning the intuitive problem: the base truths C may well include relevant truths about the distant past, including perhaps the spatiotemporal configuration of physical particles then, and so on. Even when S is an intuitively unknowable truth about the distant past, there is no corresponding intuitive problem with the idea that one can know that *if* the sentences in C are true, then S is true. Likewise there is no corresponding intuitive problem with the idea that one can know a priori a material conditional connecting a conjunction of all the truths in C to S . Something similar applies to truths about the far away and the very small. So here is no intuitive objection to the scrutability thesis here.

As for the paradox of knowability: even though Q above is unknowable, there is no formal problem with the claim that one can know that *if* the sentences in C are true, then Q is true. Indeed, as long as P itself and claims about knowledge of P are both scrutable from C , then ‘ P and no-one knows that P ’ will be straightforwardly scrutable from C . This goes for both A Priori and Conditional Scrutability.

One might suggest that the Scrutability Thesis entails the Knowability Thesis, at least if we grant that the conjunction of all truths in C is itself knowable. By knowing this conjunction D (empirically) and by knowing $D \rightarrow Q$ (a priori), one could thereby come to know Q . However, there is no reason to believe that D is itself knowable. In fact, there is good reason to believe that it is not, both for intuitive and Fitch-style reasons. The intuitive reasons are obvious: D may involve information about the distant past and the far away that no-one will ever know. As for the Fitch-style reasons: assuming that no-one in the actual history of the world believes D , then D specifies a world in which no-one believes D . If someone came to believe D , they would live in a world quite different from ours, one in which their belief would be false. So no-one can know D .

One might think that one can define a factive operator “scry” such that one scries P iff one derives P from base truths. One might then try to generate a Fitchian paradox, by taking P to be any truth that one does not actually scry, and taking Q to be ‘ P and I do not scry that P ’. By Fitch’s reasoning, if scrying is factive, then Q is an inscrutable truth. However: the notion of scrying above is ambiguous. If to scry P is to derive P from C , where C are the base truths of the actual world (or of any specific world) then scrying is not factive: there will be worlds in which P is scried but false. If to scry P is to derive P from the base truths of the world one is in at the time of scrying, then ‘ P and I do not scry that P ’ is indeed inscrutable. But this does not yield a counterexample to the A Priori or Conditional Scrutability theses above, as these theses require only that truths be derivable from the base truths of the worlds in which they are true, not the worlds in which they are so derived.

It may be that scrutability theses can do some of the work that knowability theses have been intended to do, or that they capture some of the intuitions that have led theorists to express sympathy with the knowability thesis. For example, Edgington (1985) suggests that it is intuitive that if P is true in the actual world, then it is possible that one can come to know, in some different world, that P is true in the actual world. Of course this raises questions about what it is to know

²³“Scry” is the preferred verb form of “scrutable”, having the advantage of both being more euphonious than the unlvely term “scrute” and already being a word of English with a somewhat appropriate meaning. “Scry: to divine, esp. by crystal gazing.” (Collins English Dictionary).

in a different world that P is true in the actual world. One suggestion is that to do this requires specifying the actual world with a canonical sentence D , and coming to know that if D were the case, P would be the case. Transposing this counterfactual claim into an epistemic mode (if D is the case, then P is the case), the resulting claim is not too far from the conditional scrutability thesis.

Another problem for the knowability thesis concerns cases of indeterminacy. (This problem is raised by Hawthorne (2005) for the case of omniscient knowers, but the problem generalizes.) Suppose that 42 is a borderline case of a small number, and let S be ‘42 is a big number’. On most views of vagueness, S is neither determinately true nor determinately false. On some such views, the statement $S \vee \neg S$ will be true all the same. If so, one could reason disjunctively: if S , then S is true, so S is knowable; if $\neg S$, then $\neg S$ is true, so $\neg S$ is knowable. So either S is knowable or $\neg S$ is knowable. But if S is (necessarily) indeterminate, this conclusion is implausible. One can raise a parallel problem for the scrutability thesis, yielding the conclusion that for all S , either S is scrutable (from a relevant D) or $\neg S$ is scrutable. Once again, this conclusion is implausible when S is indeterminate.

One could resist this conclusion by rejecting the law of the excluded middle and refusing to accept that $S \vee \neg S$ is true when S is indeterminate, or by holding that when S is indeterminate, it is likewise indeterminate whether S is scrutable. But perhaps the most straightforward way to avoid the problem is to understand the scrutability thesis as applying to *determinate* truths. That is, the thesis will say that when S is *determinately* true, or when $\text{det}(S)$ is true, then D implies S . On the relevant sort of view, the disjunction $\text{det}(S) \vee \text{det}(\neg S)$ will not be true in cases of indeterminacy, so the problem here will be avoided.

One might worry about cases of higher-order indeterminacy, where it is indeterminate whether $\text{det}(S)$ or $\neg\text{det}(S)$. In such a case, the best thing to say is that it is indeterminate whether S is scrutable from D . Given the presence of vagueness in language, one should expect that scrutability can be vague too. On this view, implication by base truths goes along with determinacy, and vagueness of implication goes along with vagueness of determinacy.²⁴

²⁴This view is analogous to the view that knowability goes along with determinacy, and vagueness of knowability goes along with vagueness of determinacy, suggested on behalf of the supervaluationist by Hawthorne 2005. There is an alternative view (Dorr 2003) on which the vagueness of knowability goes along with the vagueness of truth rather than the vagueness of indeterminacy. Transposed to the key of scrutability, this approach yields a view on which S is true iff S is scrutable and S is indeterminate iff it is indeterminate whether S is scrutable. If we accept the law of the excluded middle, this view will most naturally be combined with a view on which it is always the case that either S is scrutable or $\neg S$ is scrutable (cases apparently in the middle will be borderline cases of each).

One can extend the scrutability thesis to the thesis that for all S , the truth-value of S is scrutable from D , whatever this truth-value may be. To obtain the extended thesis, one could simply apply the original thesis to the statement ' S has truth-value T ', or better, one could apply the thesis to a statement such as ' $\neg S$ ', ' $\text{indet}(S)$ ', and other statements which are true iff S has a relevant truth-value. As in the cases above, then if one adopts the view of indeterminacy outlined above, these statements will be scrutable only when they are determinately true. So, for example, the claim will be that if $\text{indet}(S)$ is determinately true, then D implies $\text{indet}(S)$.

A final worry related to these matters arises from cases analogous to the Liar paradox. Say that S is 'This sentence is not scrutable from D '. Then if S is true, it is inscrutable, and if S is false, it is scrutable. Either way we have a counterexample to the thesis that a sentence is true if and only if it is scrutable.

This worry is an instance of a general worry for any thesis holding that a sentence is true iff it has property ϕ . Whether 'This sentence does not have ϕ ' is true or false, it generates a counterexample to the thesis. I do not think it is reasonable to infer that no such thesis can be true. If this were correct, the Liar Paradox would generate a counterexample to 'Every sentence is true iff it is true'. Instead, it seems best to say that sentences like 'This sentence does not have ϕ ' should be handled by whatever mechanism best handles the Liar Paradox. Indeed, one might take it to be a constraint on solutions to the Liar Paradox that they should also apply to sentences like this.

The most obvious thing to say is that in cases like this, ' S does not have ϕ ' is indeterminate. Given the discussion above, 'This sentence is not scrutable from D ' is slightly more complicated, as the relevant thesis says that a sentence is *determinately* true iff it is scrutable. This renders the sentence at issue more closely analogous to the Strengthened Liar, 'This sentence is not determinately true'. So a proponent of the Scrutability Thesis should say that the sentence has the same truth-value of the Strengthened Liar, whatever that truth-value is (perhaps involving some sort of higher-order indeterminacy). Saying more requires an adequate treatment of Liar paradoxes in general, but that is a problem for everyone, and not for the scrutability thesis in particular.

Finally, the scrutability thesis is in some limited respects reminiscent of the logical empiricists' verification principle, which says that only verifiable statements are meaningful. The scrutability thesis, rephrased, says that only scrutable statements are true, where a statement is scrutable if it is implied by certain base statements. Perhaps scrutability here might be seen as a sort of idealized verifiability, conditional on those statements in the base. One might then wonder whether any of the famous problems for the verification principle will apply here.

Most traditional worries about verifiability are removed by the extension of the base. Scrutability is much weaker than verifiability, not least because the base statements may include truths that are not themselves verifiable. For example, they may include truths about the distant past, the far away, about other minds, and about the extent of the universe. Because of this, there is no problem for scrutability generated by distinct empirically equivalent theories in physics, for example, or by statements about the past, or by the possibility of unverifiable ghosts.

Another famous problem is: is the verification principle itself verifiable? One might likewise ask: is the scrutability thesis itself scrutable? I will argue later that certain general versions of the scrutability thesis are themselves a priori, and are therefore scrutable. Other versions, such as scrutability from a specific base, are a posteriori. But we will later see that as long as a sort of “that’s-all” sentence is included in the base, the scrutability thesis itself will follow. In some cases this “that’s-all” sentence will itself be a sort of scrutability thesis, but this just brings out a way in which the scrutability thesis is far more flexible than the verification principle.

It is also worth noting that where the logical empiricists offered the verification principle in a prescriptive spirit, I am not inclined to offer the scrutability thesis in this way. Instead, in the first instance I am simply arguing for its truth. Perhaps downstream from these arguments, it can be used prescriptively, as a check on realism about certain subject matters that are not scrutable from base truths. Much here will depend on what one antecedently allows into the base, so the matter is not cut and dried. But in any case, it seems clear that the standard reasons for doubt about the verification principle do not apply to the scrutability thesis.

Second Excursus: The Inscrutability of Reference and the Scrutability of Truth

Quine's thesis of the Inscrutability of Reference says roughly that there is often no fact of the matter about what an expression refers to, because there are too many equally good candidates. This thesis is a metaphysical rather than an epistemological thesis: it concerns the existence and determinacy of reference, rather than our knowledge of reference. Because of this, and because the term "inscrutability" suggests an epistemological thesis, Quine later came to think that this name for the thesis was inappropriate, and renamed it the thesis of the *indeterminacy of reference*.

My scrutability thesis, unlike Quine's, is an epistemological thesis. Still, there are epistemological theses in the vicinity of Quine's thesis. In particular, Quine's metaphysical thesis of the indeterminacy of reference can be seen as a challenge to an epistemological thesis about reference. If we start from this thesis, and modify it to meet Quine's challenge and other challenges, this leads us to something like the scrutability theses I have discussed.

To start with: if we follow Quine's suggestion and reserve "scrutability" for broadly epistemological theses, one might call the following thesis the *scrutability of reference*.

The Scrutability of Reference: For any referring expression *E*, once we know enough about the world, we are in a position to know what *E* refers to.

The thesis has commonsense appeal. At the beginning of enquiry, we may not know what a term such as 'Hesperus', or 'Jack the Ripper', or 'arthritis' refers to. But once we discover enough about the world—which heavenly bodies are where, who murdered whom, which diseases have which properties—then we are in a position to know the referents of these terms. I have suggested a picture like this informally in section 3 above.

Still, there are various potential problems with the thesis as it stands. One problem is that it is not clear just what counts as knowing what an expression refers to. One might suggest that to know what 'Hesperus' refers to, one must know, of some object, that 'Hesperus' refers to it. But it is notoriously hard to give a precise content to the notion of (*de re*) knowledge of an object. Arguably, one expresses *de re* knowledge by saying "'Hesperus' refers to *that*", looking in the sky, or perhaps even by saying "'Hesperus' refers to Hesperus". But this sort of knowledge is much easier to obtain than the more substantial knowledge of reference envisaged in the scrutability thesis, such as the knowledge that we have after we do some astronomy. Alternatively, one might suggest that to know what 'Hesperus' refers to, one must have knowledge expressible in the form

“Hesperus refers to X ”, where X is a special sort of canonical designator. But here it is not clear what to count as a canonical designator of an object. For example, if ‘Venus’ is a canonical designator, does this mean that any user of ‘Venus’ knows what ‘Venus’ refers to?

Another problem is Quine’s indeterminacy thesis. If reference is indeterminate, so that there is no fact of the matter about what our expressions refer to, then we cannot know what our expressions refer to. Quine argues that there are multiple ways of assigning referents to our terms that make sense of all available data (including data about our judgments concerning whether sentences containing those terms are true), and that there is no fact of the matter about which assignment is correct. Even if one has doubts about the generality of Quine’s argument, many have made similar arguments concerning specific domains. For example, Benacerraf (1965) argues that many different sorts of entities are equally well-qualified to be the referents of number terms, all yielding the same truth-values for numerical statements. In a related way, Lewis (1993) argues that we can take various different entities to be the referent of ‘cat’, while Horgan (1986) argues that we can take various different sorts of entities to be the referent of ‘symphony’. In all of these domains, some have argued that reference is indeterminate.

Strikingly, both of these problems can be bypassed if we move from the scrutability of reference to the scrutability of truth.

The Scrutability of Truth (informal version):²⁵ For any truth S , once we know enough about the world, we are in a position to know that S is true.

The scrutability of truth captures much of the force of the scrutability of reference. The former thesis implies that for *any* true claim of the form ‘Hesperus is X ’, then once we know enough about the world, we are in a position to know that ‘Hesperus is X ’ is true. So we are in a position to know the truth-value of ‘Hesperus is Venus’, ‘Hesperus is the second planet from the Sun’, and so on for any designator at all. Most of the intuitive backing behind the scrutability of reference (e.g. that given enough qualitative information, we can know who Jack the Ripper is) is reflected in the scrutability of truth (e.g. that given enough qualitative information, we can know whether Jack the Ripper was Prince Albert Victor).

In the reverse direction, it is arguable that the scrutability of reference entails the scrutability of truth. If one holds with Frege that sentences are referring expressions that refer to their truth-values, then the entailment is immediate. Even if one rejects this claim, it is not hard to construct

²⁵For a perfect analogy with “The scrutability of reference”, this should really be “The scrutability of truth-value”. But “the scrutability of truth” serves just as well.

a referring expression that functions to refer to the truth-value of a given sentence: we might just use ‘the truth-value of S’, or perhaps better (in order to avoid semantic ascent), we might stipulate an operator ‘whether’ such that ‘whether S’ behaves this way. Then applying the scrutability of reference to these expressions yield the scrutability of truth.

Furthermore, the first problem concerning knowledge of reference has no parallel in the case of knowledge of truth. Truth is canonically presented under the concept *true*. To know that *E* is true, it suffices to have knowledge of the form *E is true*, deploying this concept. Further, if one simply knows the truth of ‘Hesperus is X’ for any relevant X, then it seems reasonable to say that one knows what Hesperus refers to.

Importantly, Quine’s central case for the inscrutability of reference causes no problems for the scrutability of truth. This case starts by assuming that the truth-values of sentences are fixed, and makes the case that there are multiple assignments of reference that yield the same truth-values. Even if this argument makes a case for the indeterminacy of reference, it does not make a corresponding case for the indeterminacy of truth: while reference varies between the multiple assignments, truth-values do not. One might suggest that if reference is indeterminate, truth must then be indeterminate too, but this is far from obvious: if one accepts Quine’s picture here, one will presumably accept a picture on which determinate truth-values do not require determinate referents (perhaps denying that truth-value is determined by referents, or perhaps holding that truth-value is determined by supervalueing over possible assignments of reference). In any case, there is certainly no direct argument for the indeterminacy of truth-value here.

Quine has other arguments for the indeterminacy of truth-value, tied to his arguments for the indeterminacy of translation. These arguments do not start by holding fixed the truth-value of all sentences, but only the truth-value of certain observational sentences. In this case, multiple assignments of reference are put forward in a way that makes a difference to the truth-value of non-observational sentences. This is a case for the indeterminacy of reference that also makes a case for the indeterminacy of truth-value. But these arguments concerning the indeterminacy of translation are usually held to be distinct from the central arguments concerning the inscrutability of reference. Further, these arguments are often held to be significantly weaker than the arguments concerning the inscrutability of reference, because they rest on much stronger verificationist or behaviorist assumptions. If this is right, Quine’s best case for the indeterminacy of reference does not undermine the scrutability of truth.

As for related arguments, such as Benacerraf’s, these have at best minor implications for matters concerning truth. In these arguments, as with Quine’s, the multiple assignments of reference

are usually chosen precisely so that they preserve the truth-values of first-order sentences (such as '2+2=4' and 'There are an infinite number of primes') in the domain in question. If so, almost all of the indeterminacies will drop out when it comes to the truth-values of statements. An exception may be quasi-philosophical statements as 'the number two is a set of sets', and the like. But now the issue is restricted to a few isolated sentences in the metaphysical domain, and these can be handled in the same way that one handles other sentences with indeterminate truth-value. The highly limited indeterminacy here contrasts with the issue concerning reference, which potentially affects every use of the relevant words, thereby rendering the scrutability of reference either false or useless.

The moral is that the inscrutability of reference is quite compatible with the scrutability of truth. Even if one is inclined to accept the arguments for the inscrutability of reference (I am not), one does not have corresponding reasons to reject the scrutability of truth.

Of course the thesis of the scrutability of truth is still informal and unclear in certain respects. We need to clarify "know enough about the world", for example so that this does not typically allow the trivializing knowledge that *S* is true. The obvious thought is that the relevant information about the world should be restricted to a limited (compact) vocabulary, and that the relevant class of truths is limited (compact) in a similar way.

So clarified, the thesis now becomes:

The Scrutability of Truth (second version): There is a compact class *C* of truths such that for all truths *S*, once we know enough truths in *C*, we are in a position to know that *S* is true.

This version of the thesis is quite close to the Inferential Scrutability thesis. It is subject to the Fitch-style problems discussed in the first excursus, but as discussed there, one can get around these by changing the scope and using a conditional formulation. This yields a version of Conditional Scrutability: there is a compact class of truths such that we are in a position to know that *if* these truths are true, then *S* is true.

This line of motivation does not yet get to A Priori Scrutability, but one might get there by a certain clarification of "in a position to know". One natural thought is that being in a position to know such-and-such should involve being able to know such-and-such, given ideal rational reflection and without further empirical information. Furthermore, it is natural to interpret the second version of the thesis above as holding that the compact class of base truths contains all the empirical information that is required to know the truth in question. In the conditional version

of the thesis, all this information is built into the antecedent of the conditional. So it is natural to require that this conditional can be known (on ideal rational reflection) without any further empirical information at all; that is, that it can be known a priori. This yields the following.

Scrutability of Truth (final version): There is a compact class of truths such that for all truths S, there is a conjunction D of truths in this class such that 'If D, then S' is knowable a priori.

This is a version of the A Priori Scrutability thesis. In this fashion, A Priori Scrutability can be motivated by starting from claims about the scrutability of reference and by modifying them to avoid the most pressing objections.