1

Scrutability and the *Aufbau*

1 Primitive concepts

What are the basic elements of thought? It is common to hold that thoughts, such as *Galahs are pink*, are composed of concepts, such as *galah* and *pink*. It is also common to hold that many concepts are composed from 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*.

In his manuscript 'De Alphabeto Cogitationum Humanarum', Leibniz suggests that there is a level of concepts so simple that they make up an alphabet from which all thoughts can be composed:

> The alphabet of human thoughts is a catalog of primitive concepts, that is, of those things that we cannot reduce to any clearer definitions.[1]

In *An Essay Concerning Human Understanding*, John Locke develops such a picture. He introduces complex ideas (or concepts) as follows:

> As simple ideas are observed to exist in several combinations united together, so the mind has a power to consider several of them united together as one idea; and that not only as they are united in external objects, but as itself has joined them together. Ideas thus made up of several simple ones put together, I call complex;—such as are beauty, gratitude, a man, an army, the universe. (Locke 1690, book 2, chapter 12)

Locke held that all of our perception and thought derives from simple ideas. At one point in the *Essay* (book 2, chapter 21), he suggests that the most basic ideas come down to eight. Three are ideas of matter that come to us through our senses: *extension*, *solidity*, and *mobility* (the power of being moved). Two are ideas

[1] Translated from Leibniz's 'De Alphabeto Cogitationum Humanarum' (A 6.4.270), written around 1679–81. Thanks to Brandon Look for the translation.

OUP UNCORRECTED PROOF – REVISES, 08/01/12, SPi

of mind that come to us through reflection: *perceptivity* (the power of perception or thinking) and *motivity* (the power of moving). The last three are neutral ideas that come to us both ways: *existence*, *duration*, and *number*.

The same theme can be found 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 by her 2009 book *Experience, Evidence, and Sense* these had expanded to the following list of 63 primes.

Substantives: *I*, *you*, *someone*, *something/thing*, *people*, *body*.
Relation substantives: *kind, part.*
Determiners: *this*, *the same*, *other/else*.
Quantifiers: *one*, *two*, *some*, *all*, *much/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, possession: *to be* (*somewhere*), *there is*, *to have*, *to be* (*someone/something*).
Life and death: *live*, *die*.
Time: *when/time*, *now*, *before*, *after*, *a long time*, *a short time*, *for some time*, *in a moment*.
Space: *where/place*, *here*, *above*, *below*, *far*, *near*, *side*, *inside*.
Logic: *not*, *maybe*, *can*, *because*, *if*.
Augmentors: *very*, *more*.
Similarity: *like.*

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 (from Goddard 2003, p. 408) 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.

In twentieth-century philosophy, this sort of framework was developed most systematically by Bertrand Russell and Rudolf Carnap.[2] Russell suggested that

[2] Russell engaged in numerous different projects of analysis and construction. Some central works concerning analysis into primitives involving acquaintance include 'Knowledge by Acquaintance and Knowledge by Description' (1911), *The Problems of Philosophy* (1912), and 'Theory of Knowledge' (1913). He pursued related projects of constructing the world from primitives

0001552230.INDD 2 8/1/2012 10:13:08 PM

OUP UNCORRECTED PROOF – REVISES, 08/01/12, SPi

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. Carnap's primitive concept was a concept of the relation of phenomenal similarity: similarity in some respect between total experiences (roughly, momentary slices of a stream of consciousness) had by a subject at different times.[3] For example, if a subject has two experiences both involving a certain shade of red, the experiences will stand in this relation of similarity. Using this simple concept, Carnap gave 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.

In Carnap's framework, these concepts are used to build up all of our concepts of 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.[4]

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 we can 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

in numerous later works, such as 'The Philosophy of Logical Atomism' (1918). Also worth mentioning is Ludwig Wittgenstein's conception of the world as the totality of atomic facts in his *Tractatus Logico-Philosophicus* (1921), although Wittgenstein says less than Carnap and Russell about the character of his primitives and about the construction of ordinary concepts.

[3] In this book 'phenomenal' always means 'experiential': roughly, pertaining to conscious experiences.

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

0001552230.INDD 3 8/1/2012 10:13:08 PM

OUP UNCORRECTED PROOF – REVISES, 08/01/12, SPi

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 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. We then need to say what a definition statement is, and what it is for such a statement to be adequate.

A definition statement connects a left-hand side involving a defined expression *E* to a right-hand side, with a logical form that depends on the grammatical category of *E*. Various different logical forms might be required, 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 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: 'all bachelors are untidy men' is not true in virtue of meaning, one cannot know a priori that all bachelors are untidy men, and it is not true in all possible worlds that all bachelors are untidy men. 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. Carnap intended the *Aufbau* to shed light on knowledge and on meaning, but it is questionable whether definitions that are merely extensionally adequate can fulfill these epistemological

0001552230.INDD 4 8/1/2012 10:13:08 PM

OUP UNCORRECTED PROOF – REVISES, 08/01/12, SPi

and semantic goals. For example, while a definition of 'bachelor' as 'unmarried man' may shed some light on the meaning of 'bachelor' and on how we come to knows 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.[5]

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.

Definitions allow us to connect sentences in different vocabularies. Given a definition of bachelors as unmarried men, 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,[6] 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 from which all truths are definitionally scrutable.

[5] Carnap sometimes explicitly invokes stronger criteria in the *Aufbau*. For example (as Chris Pincock pointed out to me), in section 49 he suggests a method according to which constructional definitions for scientific objects are determined by their epistemological 'indicators'.

[6] We 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 this substitution will not change truth-values in the language. 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 *x*, *x* believes that such-and-such if…').

0001552230.INDD 5 8/1/2012 10:13:08 PM

OUP UNCORRECTED PROOF – REVISES, 08/01/12, SPi

Here, a truth is a true sentence.[7] 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* that definitionally entails all truths. The world-sentence says that there exist entities that are related in such-and-such fashion by the phenomenal similarity relation *R*. If there are just two dissimilar total experiences in the world, then the world-sentence will be a sentence saying that there are two entities that stand in *R* to themselves but not to each other: $\exists x, y\,(Rxx \,\&\, Ryy \,\&\, {\sim}Rxy \,\&\, {\sim}Ryx \,\&\, \forall w(w = x \vee w = y) \,\&\, {\sim}(x = y))$. 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.[8] 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 from which all truths are a priori scrutable.[9]

[7] The choice of sentences rather than propositions here is discussed in 2.2. A subtlety here (discussed at length in the third excursus) 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, we can talk instead of the scrutability of sentences in contexts.

[8] For the world-sentence just specified, *R* will be defined as that relation *R′* such that: $\exists x, y\,(R'xx \,\&\, R'yy \,\&\, {\sim}R'xy \,\&\, {\sim}R'yx \,\&\, {\sim}\forall w(w = x \vee w = y) \,\&\, {\sim}(x = y))$. Then the new world-sentence will be the resulting of replacing *R* everywhere in the world-sentence above by this definition. Or more straightforwardly, the world-sentence can simply say $\exists R', x, y\,(R'xx \,\&\, R'yy \,\&\, {\sim}R'xy \,\&\, {\sim}R'yx \,\&\, {\sim}\forall w(w = x \vee w = y) \,\&\, {\sim}(x = y))$.

0001552230.INDD 6 8/1/2012 10:13:08 PM

OUP UNCORRECTED PROOF – REVISES, 08/01/12, SPi

We can define a priori scrutability in parallel to definitional entailment: a sentence *S* is a priori scrutable from (or a priori entailed by) a class of sentences *C* if *S* can be logically derived from some members of *C* along with some a priori truths. Given weak assumptions,[10] 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*' (which is equivalent to '$\sim(D \,\&\, \sim S)$') 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.

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 also 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

[9] A 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.

[10] In 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.

0001552230.INDD 7 8/1/2012 10:13:11 PM

OUP UNCORRECTED PROOF – REVISES, 08/01/12, SPi

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 spacetime 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 gives a general argument against phenomenalism: the view that statements about the external world can be definitionally analyzed in purely phenomenal terms. On a phenomenalist view, '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 $\sim R$. If so, no phenomenalist analysis of P can succeed.

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*.[11] 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 co-instantiation. Newman and Carnap observe that as long as we are liberal enough about what we count as a property or a relation, this world-sentence will be satisfied almost vacuously.[12] Carnap responds by suggesting that the properties and relations in question must be

[11] Carnap marks these sections of the *Aufbau* (153–55) '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.

0001552230.INDD 8 8/1/2012 10:13:11 PM

OUP UNCORRECTED PROOF – REVISES, 08/01/12, SPi

restricted to 'natural' (or 'founded', or 'experiencable') 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.[13]

One might wonder whether expanding the base like this is in the spirit of the *Aufbau*. For many years, the popular conception of logical empiricism has focused on a commitment to phenomenalism and verificationism (views on which a phenomenal base is central), and the *Aufbau* has been regarded as a paradigm of that tradition.[14] In reality, these views do not play a central role in the *Aufbau*. A much more important role is played by Carnap's commitment to structuralism and objectivity in developing a language for science. 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. A base with expressions for specific sensory qualities or specific physical properties (such as

[12] In 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.

[13] Even while retaining a phenomenal base, Carnap has some options in avoiding the first three problems. Carnap's construction is defended against Goodman and Quine by Thomas Mormann (2003, 2004), while a different construction from an expanded phenomenal base is explored by Hannes Leitgeb (2011).

[14] The distortions in the popular conception of the *Aufbau* and logical empiricism are explained partly by simplified versions promulgated by A. J. Ayer and W. V. Quine, and partly by a post-*Aufbau* period in the Vienna Circle in which phenomenal reductions involving protocol sentences played a more crucial role. Within a few years of that period (for example, in his 1932 work 'The Physical Language as the Universal Language of Science') Carnap had moved on again 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* and logical empiricism, including Alberto Coffa's *The Semantic Tradition from Kant to Carnap* (1985), Michael Friedman's *Reconsidering Logical Positivism* (1999), Alan Richardson's *Carnap's Construction of the World* (1998), and Thomas Uebel's *Overcoming Logical Positivism from Within* (1992), among other works, has painted a picture that is much more nuanced than the popular caricature.

0001552230.INDD 9 8/1/2012 10:13:11 PM

OUP UNCORRECTED PROOF – REVISES, 08/01/12, SPi

spatiotemporal properties) might not fully vindicate Carnap's structuralism, but as I discuss in chapter 8, there are other bases that come even closer to fulfilling Carnap's goals. In any case, expanded bases have the potential to fulfill many of the more general aims of a project of definability, while avoiding the criticisms above.

Other doubts about the project of the *Aufbau* are driven not by Carnap's basic vocabulary but by his construction method: that is, by his method of deriving nonbasic truths from basic truths using definitions. A number of doubts about definitions have been influential.

First: In 'Verifiability' (1945), Friedrich Waismann argued that purported definitions of ordinary expressions are subject to the problem of *open texture*: these definitions are always subject to correction, as we cannot foresee all possibilities to which they might apply. Every definition 'stretches into an open horizon', and no definition of an empirical term will cover all possibilities. Waismann's argument was especially directed at definitions in the style of logical empiricism that appeal to methods of verification, but his underlying point applies quite generally.

Second: In the *Philosophical Investigations* (1953), Ludwig Wittgenstein suggested that when we apply a term such as 'game' to some things, there is no single condition that they all satisfy. 'Game' is a family resemblance term, with different sorts of games resembling each other in various respects and with no common core. There is merely a 'complicated network of similarities, overlapping and criss-crossing'. Many have taken this idea to suggest that there are no definitions giving necessary and sufficient conditions associated with ordinary expressions of this sort.

Third: In 'Two Dogmas of Empiricism' (1951), Quine gave a critique of the notion of definition and more generally of the analytic/synthetic distinction He argued that standard understandings of these notions are circular and that the notions are based on a misconceived picture of language and its relation to the world. 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.

Fourth: In *Naming and Necessity* (1980), Saul Kripke argued against descriptivism: the thesis that names are equivalent to descriptions. Kripke's modal argument makes a case that for an ordinary name (e.g. 'Aristotle') and an 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 an ordinary name (e.g. 'Gödel') and an associated description (e.g. 'the man who proved the incompleteness of arithmetic'), the name and the description are not a priori equivalent. If these arguments succeed, then it appears that no *Aufbau*-

0001552230.INDD 10 8/1/2012 10:13:11 PM

OUP UNCORRECTED PROOF – REVISES, 08/01/12, SPi

like definitional project that applies to names and that invokes necessity or apriority as a condition of adequacy can succeed.

These criticisms mainly threaten an *Aufbau*-style project whose construction relation requires definitions of nonbasic expressions. Just as we can get around the first class of problems by expanding the base, we can get around the second class of problems by weakening the construction relation.

Before doing that, however, it is useful to look more closely at the source of the problems. At least three of the critiques (Waismann's, Wittgenstein's, and Kripke's) turn on a common problem: the problem of *counterexamples.* (Quine's critique turns on somewhat different issues, and I return to it in Chapter 5.) For many terms in English, it seems that every definition that has ever been offered is subject to counterexamples: actual or possible cases to which the original term applies but the purported definition does not, or vice versa, thereby showing that the definition is inadequate.

The most famous case is the case of 'knowledge', traditionally defined as 'justified true belief'. In his 1963 paper 'Is Knowledge Justified True Belief?', Edmund Gettier pointed out 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 of knowledge 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 (Shope's *The Analysis of Knowing* gives 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.[15] And even in the case of 'bachelor', there are unmarried men who do not seem to be bachelors, such as those in

0001552230.INDD 11 8/1/2012 10:13:11 PM

OUP UNCORRECTED PROOF – REVISES, 08/01/12, SPi

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.

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 (1975) 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.[16] 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 to a priori scrutability[17]

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

[15] A philosopher will find possible counterexamples to many or most of Wierzbicka's definitions. Wierzbicka's intended criteria of adequacy for definitions almost certainly differ from philosophers' criteria, so it is not obvious to what extent the existence of counterexamples is a problem for Wierzbicka's project.

[16] A distinct anti-definition influence in psychology derives from psycholinguistic arguments for the conclusion that lexical concepts are primitive by Jerry Fodor et al. (1980).

[17] This section overlaps in part with Chalmers and Jackson 2001.

0001552230.INDD 12 8/1/2012 10:13:12 PM

OUP UNCORRECTED PROOF – REVISES, 08/01/12, SPi

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 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 usually 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 method of cases is precisely how counterexamples to definitions are often generated. When someone suggests that *E* can be defined as *F* ('bachelor' is defined as 'unmarried man', say), someone else suggests a scenario *D* (involving long-term gay couples, say) 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 ~*K*, and so on.

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 ~*K*' a priori. Someone who knows that *G* is true and who has mastered 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, mastery of the concept of knowledge (along with a grasp of

0001552230.INDD 13 8/1/2012 10:13:12 PM

OUP UNCORRECTED PROOF – REVISES, 08/01/12, SPi

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, ~*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 a key 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' and the other terms involved. But the case above is another counterexample. At least once general skepticism about the a priori is set aside, 'If *G* then ~*K*' is a central 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 knowl-

0001552230.INDD 14 8/1/2012 10:13:12 PM

OUP UNCORRECTED PROOF – REVISES, 08/01/12, SPi

edge. 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. Rather, at least in common practice, proposed sufficient conditions are hostage to a priori intuitions about specific cases.

It may even be that there are no short nontrivial sufficient conditions for knowledge. That is, it may be that any reasonably short condition not involving 'know' or cognates is compatible with the absence of knowledge.[18] Not every expression is like this. For example, there are plausibly short sufficient conditions for *not* knowing that *p*: the condition of not believing that *p*, or of believing that *p* based solely on inference from a false belief. But it may be that for many expressions, there are at least hypothetical cases for which there is no reasonably short nontrivial sufficient condition (perhaps even no finite sufficient condition) obtaining in that case. In such a case, a nontrivial sufficient condition must be a long one: in the limit, a fully detailed specification of such a scenario, perhaps in the language of a scrutability base. All this is quite consistent with A Priori Scrutability, but it does bring out the need for idealization in understanding the thesis.

An opponent of A Priori Scrutability may hold that there are not even long nontrivial sufficient conditions for knowledge and the like, or that any sufficient conditions here do not yield a priori scrutability. These remain separate substantive issues, distinct from the standard objections to Definability and addressed in the arguments for A Priori Scrutability in later chapters. For present purposes, it suffices to observe that the standard objections to Definability are not objec-

[18] See 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 deviant cases compatible with these descriptions in which subjects have other evidence for the relevant *p* (see Malmgren 2011 and Ichikawa and Jarvis 2009 for discussion). *G* above may escape this charge by including the 'based solely on' clause. But the point still applies to justification: there will be deviant possible cases that satisfy *G* but not *J* because extraneous factors undermine Smith's justification for believing the relevant proposition. Deviant cases undermine conclusive a priori scrutability (in the sense of 2.1) of *J* from *G* and may undermine any short nontrivial a priori sufficient condition for justification, but they do little to undermine the weaker scrutability claim that *J* is nontrivially a priori scrutable from a full enough specification of the case. An analogy: deviant cases undermine necessitation of *J* by *G* and may undermine any short nontrivial modally sufficient condition for justification, but they do little to undermine the weaker supervenience-style claim that *J* is nontrivially necessitated by a full enough specification of the case.

0001552230.INDD 15 8/1/2012 10:13:12 PM

OUP UNCORRECTED PROOF – REVISES, 08/01/12, SPi

tions to A Priori Scrutability and that A Priori Scrutability remains an attractive thesis in the face of them.

4 From descriptions to intensions[19]

At this point we can take a leaf from Carnap's later work, especially his 1947 book *Meaning and Necessity*, and understand the meaning of expressions not in terms of definitions but in terms of *intensions*. Here the intuitive idea is that an intension captures the way an expression applies to possible cases of all sorts. For example, the Gettier case brings out that whether or not there is a good definition for 'know', we can classify different scenarios as involving knowledge or as not involving knowledge. An intension is a way to represent those classifications.

The intension of an expression can be identified with a function from scenarios to extensions, mirroring speakers' idealized judgments about the extension of the expression in the scenario. The intension of a sentence (as used in a context) is a function from scenarios to truth-values. For example, the intension of 'Smith knows that Jones owns a Ford or Brown is in Barcelona' is false in a Gettier scenario. The intension of a subsentential expression such as 'bachelor' is a function from scenarios to sets of individuals. In any given scenario, its intension picks out the people who are bachelors if that scenario is actual. An expression's intension will often depend on its context of use, but for simplicity I will set aside this context-dependence for now.

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. (Here and throughout, I work with an idealized notion of epistemic possibility that is tied to what cannot be ruled out a priori.) For a given scenario *w* and a given sentence *S*, we can consider the hypothesis that *w* actually obtains and judge whether if *w* obtains, *S* is the case. If yes, the intension of *S* is true at *w*. If no, the intension of *S* is false at *w*. I give a fuller definition of scenarios and intensions in the tenth excursus, but for now we can work with this intuitive understanding.

On this model, speakers can grasp an expression's intension without grasping a corresponding definition. Instead, the grasp corresponds to a *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

[19] This section presupposes a little more philosophical background than the rest of the chapter and can be skipped without too much loss by nonspecialists. There is a somewhat gentler introduction to the framework of intensions in chapter 5, sections 3–5. A more precise account is in the tenth excursus.

0001552230.INDD 16 8/1/2012 10:13:12 PM

OUP UNCORRECTED PROOF – REVISES, 08/01/12, SPi

subject using expressions such as 'bachelor', 'knowledge', and 'water' will have the ability to evaluate certain conditionals of the form 'If *E*, then *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 saying whether a given case fall into its extension (e.g. 'John is a bachelor', 'Sue knows that *p*', 'Water is H_2O'). 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.

For some expressions, we can capture the intension of the expression in the form of a definition. In other cases, we will merely be able to approximate an intension with an *approximate definition*. For example, 'justified true belief' can be seen as an 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. If there is no finite definition that gets all possible cases right, there may be a converging series of definitions: a series of longer and longer approximate definitions such that for any given case, there is some point in the series after which all definitions get that case right. In all these cases, however, the definitions are beholden to the intension rather than vice versa.

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 on the intensional model, 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).

All this applies equally to Kripke's arguments against descriptivism, which are also arguments from counterexample. In fact, Kripke deploys two different sorts of arguments from counterexample. We might say that *modal* arguments from counterexamples are used to oppose the claim that '*N* is the *D*' is necessary (for a name *N* and a description *D*), while *epistemic* arguments from counterexample are used to oppose the claim that '*N* is the *D*' is a priori. In the case of knowledge, the Gettier counterexample serves as the basis of both a modal argument and an epistemic argument, showing that it is neither necessary nor a priori that knowledge is justified true belief. In Kripke's arguments against descriptivism, modal arguments and epistemic arguments from counterexample are employed separately.

0001552230.INDD 17 8/1/2012 10:13:12 PM

OUP UNCORRECTED PROOF – REVISES, 08/01/12, SPi

Modal arguments from counterexample require exhibiting a *metaphysically possible* situation (roughly, a situation that might have obtained) of which the equivalence is false. Kripke's modal argument against descriptivism fits this template. 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 it would not have been the case that Aristotle was the teacher of Alexander. It follows that it is not necessary that Aristotle was the teacher of Alexander.

Epistemic arguments from counterexample require exhibiting an *epistemically possible* scenario (that is, a scenario not ruled out a priori) of which the equivalence is false. Kripke's epistemological argument against descriptivism is an argument from counterexample of this 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, modal arguments from counterexample show 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. As usual, we need only choose an intension that respects the counterexample. The modal profile of 'know' can be represented as an intension that classifies Gettier cases as cases in which knowledge is absent. Likewise, the modal profile of 'Aristotle' can be represented as an intension that picks out Aristotle in the situation in which he never went into pedagogy, rather than picking out Alexander's teacher.

Similarly, epistemic arguments from counterexample show 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.[20] The epistemic profile of 'knows that P' can be represented as an intension that classifies Gettier cases as cases in which knowledge is absent. Likewise, the epistemic profile of 'Gödel' can be represented as an intension that picks out the stealer in Kripke's stolen-proof scenario rather than the prover.

[20] In the case of an expression such as 'knowledge', the epistemic and modal 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 epistemic and modal 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). The intension over epistemically possible scenarios discussed in the text is the primary intension, which is the most important for present purposes.

0001552230.INDD 18 8/1/2012 10:13:12 PM

OUP UNCORRECTED PROOF – REVISES, 08/01/12, SPi

Like Gettier's argument from counterexample, Kripke's arguments from counterexample pose no problem for A Priori Scrutability. Kripke's modal argument does not pose even a prima facie problem: it concerns what is metaphysically possible and necessary, whereas A Priori Scrutability concerns what is a priori and a posteriori. Kripke's epistemological argument suggests that 'Gödel' is not a priori equivalent to a description such as 'the prover of incompleteness', but it gives no reason to deny that sentences such as 'Gödel did not prove incompleteness' are themselves scrutable from a specification of the relevant scenario. Given a specification of the stolen-proof scenario, we can certainly determine that if the scenario is actual, Gödel did not prove incompleteness.

Likewise, Kripke's epistemological argument cannot refute *approximate descriptivism*: the thesis that for every name (as used by a speaker) there is a converging series of descriptions such that for every scenario, there is some point in the series such that all descriptions after that point give the same result as the name in that scenario. An approximate definition that works fairly well for 'Gödel' is 'The actual person called 'Gödel' by those from whom I acquired the name'.[21] As usual the approximation will be imperfect and there will be counterexamples (cases where one misheard the name, perhaps), but refinements will gradually remove the counterexamples as they converge on the name's intension. In any case, these counterexamples pose no more of a problem for A Priori Scrutability or for the intensional model than the Gettier case.

Much follows from these observations. Kripke's arguments are often thought to undermine broadly Fregean analyses of meaning and content. But we will see shortly (and in more detail in the eleventh excursus), an appropriate scrutability thesis can itself be used to support a broadly Fregean analysis of meaning and content, by defining intensions over epistemically possible scenarios. The resulting intensions can do much of the work that descriptions or Fregean senses are often held to do.

We can put things as follows. If the scrutability thesis is correct, a Fregean view of meaning and content is viable. Kripke's arguments give us no reason to reject the scrutability thesis. So Kripke's arguments should not lead us to reject a Fregean view of meaning and content. The scrutability thesis therefore suggests that Kripke's arguments are much more limited in scope than is often supposed. Of course there is more to say here, but this at least makes an initial case that the seemingly innocuous scrutability thesis may have highly significant consequences.

[21] For more on approximate descriptivism, see 8.2. For more on intensions and approximate definitions in the 'Gödel' case, see the discussion of Kripke's epistemological argument in 'On Sense and Intension'.

0001552230.INDD 19 8/1/2012 10:13:12 PM

OUP UNCORRECTED PROOF – REVISES, 08/01/12, SPi

5 The scrutability base

A *scrutability base* is a class of truths from which all truths are scrutable, for a given notion of scrutability.[22] What sort of truths might go into a scrutability base?

At the end of the *Aufbau*, Carnap embraces what we might call Logical Scrutability: the view that there is a scrutability base using only logical expressions. Some phenomenalists accept Phenomenal Scrutability, holding that there is a scrutability base using only phenomenal expressions (expressions for 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.[23] 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 working scrutability thesis is what we might call Compact Scrutability: there is a compact class of truths from which all truths are scrutable. 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 if 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 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.

[22] I will speak of sets and classes interchangeably. For some purposes it might be useful to admit classes of sentences that are too large to form a set, but for most of our purposes set-sized classes will be adequate. I discuss this issue further toward the end of E3.

[23] In principle, these views concerning a scrutability base can be combined with different scrutability relations (such as definitional or a priori scrutability), yielding such theses as Definitional Phenomenal Scrutability, A Priori Microphysical Scrutability, and so on. When the scrutability relation is not specified, a thesis involving a priori scrutability should be understood. For more on the conventions here, see 2.1.

0001552230.INDD 20 8/1/2012 10:13:12 PM

OUP UNCORRECTED PROOF – REVISES, 08/01/12, SPi

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 C is a scrutability base and no proper subclass of C is a scrutability base. (In order to ensure that C uses a minimal vocabulary, one could also require that there is no scrutability base using only a proper subclass of the expressions used in C.)

0001552230.INDD 21 8/1/2012 10:13:12 PM

OUP UNCORRECTED PROOF – REVISES, 08/01/12, SPi

Three proposals about minimal scrutability bases correspond to 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. I will argue (in chapters 3, 4, and 6) that all ordinary macroscopic truths are a priori entailed by a class *PQTI* (physics, qualia, that's-all, indexicals) that includes both truths of physics *and* phenomenal truths, as well as certain indexical truths ('I am such-and-such', 'Now is such-and-such') and a totality or 'that's-all' truth (on which more in 3.1 and E5). If so, then *PQTI* 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, indexical, and totality expressions,[24] 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.

Another principled scrutability thesis, perhaps less in the spirit of Carnap's view, is Fundamental Scrutability: the thesis that all truths are scrutable from

[24] Throughout 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.

0001552230.INDD 22 8/1/2012 10:13:12 PM

OUP UNCORRECTED PROOF – REVISES, 08/01/12, SPi

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.[25] 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.

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 the theses just mentioned, as well as a tentative case for Fundamental Scrutability. I will also connect each of these theses to philosophical applications. For the purposes of many applications, it is these specific scrutability theses rather than Compact Scrutability per se that matters. Compactness plays a role in some applications, but where it does not, it can be seen as playing a sort of guiding role en route to the specific theses, ensuring that our scrutability bases are small enough that those theses are plausible.

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.

There are many scrutability bases. 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

[25] Metaphysical 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.

0001552230.INDD 23 8/1/2012 10:13:12 PM

OUP UNCORRECTED PROOF – REVISES, 08/01/12, SPi

bases of which no proper subclass is a scrutability base and for which there is no scrutability base using only 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 *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. Carnap himself held a pluralistic view on which there are many equally privileged bases that we can choose between only on pragmatic grounds. Still, the phenomenon does suggest that the mere fact that an expression is involved in a minimal scrutability base does not suffice for the expression to express a primitive concept in an 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 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 by postulating a more fine-grained relation of conceptual or epistemological dependence among truths. Whenever one class of truths depends on another in this sense, truths involving the former will be 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).[26]

[26] This 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. We could think of the more fine-grained relation as conceptual dependence or conceptual grounding. For more on these issues, see E16.

0001552230.INDD 24 8/1/2012 10:13:12 PM

OUP UNCORRECTED PROOF – REVISES, 08/01/12, SPi

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. I discuss such fine-grained relations and their relation to scrutability later in the book (and also in the companion chapter, 'Verbal Disputes').

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

6 Reviving the *Aufbau*

If the A Priori Scrutability thesis is correct, it offers a vindication of something like the project of the *Aufbau*.[27] There are two significant differences: the very limited bases (logical and/or phenomenal) 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 framework. 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

[27] A quite different project in a similar spirit, attempting to vindicate something like the *Aufbau*, is carried out by Hannes Leitgeb in his important article 'New Life for Carnap's *Aufbau*?' (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.

0001552230.INDD 25 8/1/2012 10:13:12 PM

OUP UNCORRECTED PROOF – REVISES, 08/01/12, SPi

wrong. I will address a number of other challenges to the scrutability framework in chapters 3 and 4.

One might ask: does A Priori Scrutability have the potential to satisfy some of the ambitions of the *Aufbau*? These ambitions included an analysis of meaning and concepts, an epistemological optimism, a 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. 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 of Carnap's ambitions in the *Aufbau* and other works.[28]

1. *Knowability and skepticism*. In the *Aufbau*, Carnap used his construction to argue that there is no question whose answer is in principle unattainable by science. This is a version of the notorious Knowability Thesis in epistemology, often associated with the programs of logical empiricism and verificationism, which 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 capture at least a plausible relative of these theses, and can play some parts of the role that the knowability thesis has been used to play. Furthermore, certain scrutability theses offer a distinctive response to skepticism (E15).

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 analyzes 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 (E10). 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, and behave

[28] For more on these applications, see E1 and E15 (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 E12 (structuralism and the unity of science), and 6.5 (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.

0001552230.INDD 26 8/1/2012 10:13:12 PM

OUP UNCORRECTED PROOF – REVISES, 08/01/12, SPi

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 the eleventh excursus, 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 set of scenarios in which it is true. Then two sentences will have the same intension if and only if 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 if and only if *'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*. Carnap's philosophy was known for its anti-realism about metaphysics: many metaphysical questions do not have objective and determinate answers. With specific scrutability theses in hand, the current framework can be used to argue for realism, anti-realism, or metaphysical primitivism about a given subject matter. For example, given Fundamental Scrutability (the thesis that all truths are scrutable from fundamental truths), then if ontological 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. In the domain of ontology, one might use this method to argue for

0001552230.INDD 27 8/1/2012 10:13:12 PM

OUP UNCORRECTED PROOF – REVISES, 08/01/12, SPi

a sort of anti-realism.[29] In other domains (that of consciousness, say), one can use this method to argue for an expansion in the metaphysically fundamental truths. We can also use scrutability as a guide in various projects of conceptual metaphysics, discussed in the sixteenth excursus.

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 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, this account can be used to provide an explanation of the scrutable truths. If so, then (as I argue in E12), 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 ('Verbal Disputes') suggests a way of reducing all philosophical disagreements to disagreements over base truths.

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.

[29] This 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.

0001552230.INDD 28 8/1/2012 10:13:12 PM