1

Compositionality

1 THE AGE OF THE UNIVERSE

Natural languages are *compositional*, or *have compositional semantics:* the meaning of a meaningful complex expression is "composed" from the meanings of its immediate constituents, in a way that reflects the syntactic configuration of those constituents. For example, if we suppose that 'Mary loves John' consists syntactically in the combination of the proper name 'Mary' and the verb phrase (VP) 'loves John', then the meaning of 'Mary loves John' will be the result of composing the meaning of 'Mary' with the meaning of 'loves John'. In turn, the meaning of 'loves John' will be the result of composing the meaning of 'loves' and the meaning of 'John'. 'Mary', 'loves' and 'John' are not themselves complex (grant this for 'loves'), so their meanings are not composed from simpler ones. Instead, their meanings have to be learned individually.

A little more formally, compositionality can be understood to say that the meaning of a complex expression is a *function* of its syntax and the meanings of its immediate constituents. A function is a black box which operates on inputs to produce an output. The workings of the box are such that different inputs may produce the same output, but the same inputs never produce different outputs (thus functions are said to be *deterministic*). For instance, the function of exponentiation produces the same output, 64, for the two different inputs (i) 2 and 6 in that order, and (ii) 8 and 2 in that order. The inputs 8 and 2 always produce 64, as do the inputs 2^3 and 2. Thus the output of a correct application of exponentiation depends only on the input numbers, and is not sensitive to the way the inputs are expressed or described ('8' versus '2^3').

We will use the brackets ⟦ and ⟧ to abbreviate 'the meaning of _', so that we can write, say, '⟦loves John⟧', instead of "the meaning of 'loves John'". And let us think of a meaning-generator for a natural language as a deterministic black box. Then giving it the three inputs ⟦Mary⟧, ⟦loves John⟧, and the syntactic configuration '[Mary [loves John]]', it will produce ⟦Mary loves John⟧ (i.e., the meaning of 'Mary loves John') as output. Repeating this procedure over and over never changes the output, and if the meaning of 'MaryBeth' is the same as that of 'Mary', then the inputs ⟦MaryBeth⟧, ⟦loves John⟧, and the syntactic configuration '[MaryBeth [loves John]]', will produce the same meaning as that of 'Mary loves John' as output. In other words, granted ⟦Mary⟧ = ⟦MaryBeth⟧, compositionality guarantees that ⟦Mary loves John⟧ = ⟦MaryBeth loves John⟧.

A good illustration of the rationale for the thesis that natural languages are compositional is provided by examples such as

(1) The universe is about fourteen thousand million years old.

There will be some readers for whom (1) is a novel sentence, one they have not encountered before, either in speech or writing. But in terms of grasping what (1) says, these readers are in no worse a position than those to whom the thought occurs every day. Typical English-speakers understand (1), even if it is novel to them, because they understand its constituent words, they perceive its syntax, and they can compose its meaning from these resources.[1] If there is a significantly different and equally plausible explanation of the ability to understand novel sentences, it has yet to be discovered.

As this example suggests, compositionality can constrain what kinds of thing the meanings of expressions can be. This constraint

[1] This example suggests, correctly, I think, that if compositionality is to be explained in terms of functional dependence, not just any old function will do: the function must be tailored to supporting these capacities. Such restrictions may make available a response to the objections in (Szabó 2000:499–500) to the functional dependence account of meaning-composition.

on the nature of meanings[2] may be broken into two parts:

(i) if the phrase $e_1{}^\frown e_2$ with constituent expressions e_1 and e_2 is meaningful, the meanings of e_1 and e_2 must be of such a nature as to permit some way of being composed with one another;

(ii) composition of meanings in whatever way that (i) provides must produce the right results – if $e_1{}^\frown e_2$ is meaningful, its meaning should be the item that is produced when the meaning of e_1 is composed with the meaning of e_2.

Following word order and writing $x(y)$ for the composition of x with y, (ii) above can be stated briefly as the constraint that $[\![e_1{}^\frown e_2]\!] = [\![e_1]\!]([\![e_2]\!])$, for all expressions e_1 and e_2 such that $e_1{}^\frown e_2$ is meaningful.

As an example of what the combination of (i) and (ii) exclude, consider the phrase 'pet fish' (Osherson and Smith 1981) and the view that the meanings of (simple *and* complex) expressions which apply to individual objects are *lists of features weighted by typicality* ('t-weighted lists of features'). If this is so, then according to (i) we should be able to *compose* lists of t-weighted features into lists of t-weighted features. And according to (ii), the list of t-weighted features that is in fact the one for 'pet fish' should be the one we get by composing the list of t-weighted features for 'pet' with the one for 'fish', by whatever method of composition we have in hand from (i). In our meaning-bracket notation, this is to say that the list ⟦pet fish⟧ must be equal to the result of composing the list ⟦pet⟧ with the list ⟦fish⟧: we need ⟦pet fish⟧ = ⟦pet⟧(⟦fish⟧).

It is a non-trivial problem for the proponent of meanings as lists of t-weighted features to explain composition so that (i) and (ii) are satisfied simultaneously. Suppose that in the list for 'pet fish', the heavily weighted features are those of a typical goldfish and a typical small multi-colored tropical fish. For ⟦pet⟧ the heavily weighted features are 'four legs', 'furry/hairy', 'playful', and so on, while for ⟦fish⟧, think of the features of something like a cross between a trout and a salmon, with a bit of shark thrown in. The problem is to describe a methodical operation on the lists ⟦pet⟧ and ⟦fish⟧ that gets us a list

[2] It is extensively argued for in (Fodor and Lepore 1996; 2002b, Chs. 1–3). But in my choice of example to follow, I am respecting some replies in (Horwich 2006: Ch. 8).

⟦pet fish⟧ that has heavy weights on goldfish-features and small multi-colored tropical fish features, and handles all similar cases.

A unification operation may suggest itself, but it is far from obvious how to unify lists when they differ over the same feature. Legs seem nearly as important to the typical pet as leglessness is to the typical fish. Does this mean that it is *open* whether or not pet fish have legs, or that leggedness is *moderately* t-weighted in ⟦pet fish⟧?

An alternative is to suppose that ⟦fish⟧ already has 'pet' as a feature with moderate to low weight, and ⟦pet fish⟧ is obtained simply by increasing the weight of 'pet' in ⟦fish⟧ to the maximum (this is the core of the 'selective modification model' of Smith *et al.* 1988). But it is a standard objection to this that the weights on fish-features are not independent. If we increase the weight of 'pet' to the maximum, then in order to get ⟦pet fish⟧, the weights in ⟦fish⟧ of (for example) habitat features like 'river', 'loch' and 'ocean' must plunge, while that of 'glass bowl' should soar. So composition of t-weighted features will *follow* the (purported) facts about what the things to which 'pet fish' applies are like. And presumably one must first understand the complex phrase to identify the things whose prominent features should be the heavily weighted ones (someone who, amazed to learn that people keep fish as pets, asks 'what are pet fish like?', understands her own question). So it is viciously circular to claim that such weighted features constitute the meaning of the complex phrase.

It is not hard to see that the underlying problem is the attempt to absorb contingent *a posteriori* matters into meanings.[3] Even if we

[3] This point seems insufficiently appreciated in (Robbins 2002), which proposes to assimilate such items as lists of weighted features to Fregean senses or narrow contents (328–9). But however these latter are explained, they must not embody what is just widely-shared collateral information. Someone who thinks that pet fish are only kept in rectangular plastic acquaria thinks exactly *that* without qualification, so his failure to accord high weight to 'glass bowl' as a habitat-feature does not impugn his grasp of ⟦pet fish⟧. Instead, it merely indicates a false belief about pet fish. Quite extraordinary false beliefs are consistent with firm grasp of meaning. Certainly, as the number of apparently sincere avowals of extraordinary falsehoods about pet fish mounts in a single speaker, the hypothesis that his understanding of 'pet fish' is deviant grows in plausibility. But that this might be the best explanation of the avowals, does not justify *identifying* ⟦pet fish⟧ with the list of t-weighted features given by normal belief-avowals using 'pet fish'. It is simply that, among those who understand 'pet fish', much the same collateral information is in the air and easily acquired.

can predict from ⟦pet⟧ that rivers, lochs and oceans are unlikely habitats for pet fish, there is no particular reason to expect *bowls*. And it is entirely contingent and *a posteriori* that those who seek the companionship of fish are not predominantly piranha enthusiasts (I believe tarantulas are the most popular pet spider). So we cannot predict just from ⟦pet⟧ and ⟦fish⟧ that the weight of 'carnivorous' in ⟦pet fish⟧ is lower, or even *thought* to be lower, than in ⟦fish⟧. The weights of features in ⟦pet fish⟧ result from empirical and contingent facts about what kinds of *fish* people are generally inclined to favour as pets, and these are not necessarily predictable from, or a function of, the empirical and contingent facts about what fish are like and about what kinds of *creatures* (let's forget about rocks) people are generally inclined to favour as pets. But the meaning of 'pet fish' *is* graspable just from grasp of the meaning of 'pet', the meaning of 'fish', and the manner of their combination in 'pet fish'. So the meaning of 'pet fish' cannot be a list of t-weighted features.[4]

2 EXTENSIONALISM

What accounts of meaning *are* compatible with compositionality? The set of things to which a predicate of objects applies is called its *extension*. For example, the common noun 'cordate' has as its extension the set of creatures with a heart; the common noun 'renate' has as its extension the set of creatures with a kidney. We will use $^\vee$ for 'the extension of', so '$^\vee$cordate' means "the extension of 'cordate'", and $^\vee$cordate = {*x: x* is a cordate}. Similarly, $^\vee$renate = {*x: x* is a renate}. A traditional candidate for the meaning of an atomic predicate is its extension: ⟦cordate⟧ = $^\vee$cordate, ⟦renate⟧ = $^\vee$renate, ⟦fish⟧ = $^\vee$fish, and so on. The equation of the meaning of a predicate with its extension we refer to as *extensionalism* about predicates.

Extensionalism about predicates solves the pet-fish problem:

(2) a. ⟦pet⟧ = $^\vee$pet = {*x: x* is someone's pet}.
b. ⟦fish⟧ = $^\vee$fish = {*x: x* is a fish}.

[4] This section has merely dipped a toe into very deep water. There is a vast literature, including (Davies 1987; Evans 1981; Fodor and Lepore 2001; Higginbotham 1986; Horwich 1998, 2006; Partee 1984; Pelletier 1994; Recanati 2003; Szabó 2000).

To satisfy principle (i) on page 3, we may explain modifier-noun composition as set-intersection. So ⟦pet⟧(⟦fish⟧) will be the set that results from taking the intersection of {*x: x* is someone's pet} and {*x: x* is a fish}, which we write $^\vee$pet ∩ $^\vee$fish. $^\vee$pet ∩ $^\vee$fish is the set of things which are both someone's pet and a fish. This set is exactly the extension of 'pet fish', hence according to extensionalism about predicates, it is ⟦pet fish⟧. Therefore principle (ii) on page 3 is also satisfied: ⟦pet⟧(⟦fish⟧) = $^\vee$pet ∩ $^\vee$fish = $^\vee$(pet fish) = ⟦pet fish⟧. Thus extensionalism can accommodate this type of complex predicate.

The concept of extension is applicable to other categories of expression. We can identify the extensions of proper names with their referents and the extensions of sentences with their truth-values. For *quantified noun phrases* (QNP's), expressions such as 'every dog', 'no black cat', 'three blind mice', and so on, the extension is the set of all sets that contain the relevant quantity of items from the extension of the predicate in the QNP.[5] Thus the extension of 'every dog' is the set of all sets that include every dog; the extension of 'no black cat' is the set of all sets that include no black cats; and the extension of 'three blind mice' is the set of all sets that include three blind mice. This is the correct choice of extension for compositional calculation of the truth-values of sentences (S) that consist in a QNP followed by a verb phrase (VP), for example,

(3) $[_{S}[_{QNP}$three blind mice$]$ $[_{VP}$ran from a farmer's wife$]]$.

(3) is true iff the set *x* of things that ran from a farmer's wife is one of the sets in the extension of 'three blind mice', which is to say, iff *x* is a set that includes three blind mice. The general rule is:

(4) $^\vee[_{S}\,\text{QNP VP}] = \top$ iff $^\vee\text{VP} \in {}^\vee\text{QNP}$.

Extensionalism about sentences, about proper names, and about quantified noun phrases, are three different claims. Each is a thesis that identifies the meanings of expressions of certain classes with their extensions. Extensionalism about an entire language $\mathcal{L}$ is the thesis that identifies the meaning of *any* meaningful expression of $\mathcal{L}$ with its extension: ⟦e⟧ = $^\vee e$ for every meaningful *e* of $\mathcal{L}$.

[5] See (Barwise and Cooper 1982); for exposition, (Larson and Segal 1995, Ch. 8).

3 PROBLEMS FOR EXTENSIONALISM

Extensionalism is more plausible for some classes of expression than others. Extensionalism about proper names is not easily refuted. But extensionalism is rather implausible for whole sentences: classically, there are only two truth-values, TRUE (for short, ⊤) and FALSE (⊥), so there would only be two meanings, all true sentences would be synonymous, and all false sentences would be synonymous. Except for expressively impoverished languages, this is almost impossible to square with compositionality (see the argument below about predicates).

Extensionalism for predicates may solve the pet-fish problem, but it is doubtful that it is the *right* way of solving it. Suppose what is anyway close to being the case, that $^{\vee}$cordate = $^{\vee}$renate. In other words, suppose {*x: x* is a cordate} = {*x: x* is a renate}, that is, that these two sets have the same members. Then if extension is meaning, we have to say that 'cordate' and 'renate' mean the same. But this sounds wrong: saying that you have a heart isn't *synonymous* with saying that you have a kidney. And if ⟦cordate⟧ ≠ ⟦renate⟧ even though $^{\vee}$cordate = $^{\vee}$renate, extensionalism about predicates is refuted.

This is just an appeal to intuitions about synonymy. It would be better if we could produce a pair of examples which differ in construction only by one's having 'cordate' where the other has 'renate', yet the two examples have different truth-values. If the only difference between the examples is that one has 'cordate' where the other has 'renate', and ⟦cordate⟧ = ⟦renate⟧, then by compositionality, the two sentences must be synonymous. But it is impossible for synonymous sentences to have different truth-values – whether or not we think of the meaning of a sentence *as* its truth-value, meaning certainly *determines* truth-value (given the facts), hence same meaning guarantees same truth-value. So consider the examples in (5):

(5) a. Cordates could have evolved without kidneys.
    b. Renates could have evolved without kidneys.

'Could have' expresses broadly logical possibility, so (5a) seems uncontroversial: presumably no *logical* necessity drove evolutionary forces to select kidneys as the engineering solution to performing the functions that kidneys in fact perform. (5b), on the other hand, makes little sense, since ‘renate’ just means ‘creature with a kidney’. Apparently, then, (5a) is true, (5b) false. But since the only difference between the meanings of (5a) and (5b) is that one has ⟦cordate⟧ where the other has ⟦renate⟧, compositionality requires that ⟦(5a)⟧ = ⟦(5b)⟧ if ⟦cordate⟧ = ⟦renate⟧. By the difference in truth-value, ⟦(5a)⟧ ≠ ⟦(5b)⟧, so ⟦cordate⟧ ≠ ⟦renate⟧.

More carefully, we should only say that (5b) has a *reading* on which it differs in truth-value from (5a). By (5b) we might mean something like ‘the creatures which in fact are renates could have evolved without kidneys’, which might be true: the very same creatures, or ones of the very same type, which *actually* have kidneys, could have evolved a somewhat different array of organs. But this is not the reading intended here for (5b); its existence illustrates a *scope ambiguity* in the sentences in (5), roughly concerning which of the bare plural noun *versus* the ‘could have’ is processed first in interpreting the sentence. For (5b) we can bring out the alternatives with the following paraphrases:

(6) a. In some possible world *w*, the renates of *w* evolve in *w* without kidneys.
b. The renates of the actual world evolve in some possible world *w* without kidneys.

The term ‘possible world’ corresponds to the intuitive idea of a way things could have gone, so (6a) claims that there is a way things could have gone in which, contradictorily, renates evolve without kidneys. (6b) says that for the actual renates there is a way things could have gone in which *they* evolve without kidneys (and hence are not renates in *that* course of events). If (5b) is taken in the sense (6a), we are giving ‘could have’ wide scope, or scope over ‘renates’, but if it is taken in the sense (6b), we are giving ‘could have’ narrow scope and ‘renates’ wide scope. It is only in sense (6a) that (5b) is contradictory. So suppose we are careful to read both (5a) and (5b) so that ‘could have’ has wide scope. This gives us sense (6a) for (5b) and its counterpart for (5a). (5a) and (5b) have different truth-values on this pair of parallel readings. But if ⟦cordate⟧ = ⟦renate⟧ they should have the *same* truth-values on any parallel readings we care to consider. So our demonstration that ⟦cordate⟧ ≠ ⟦renate⟧ still stands.

4 INTENSIONALISM

The argument of Section 1.3 indicates that we should not identify the meanings of predicates with their extensions. However, example (5) suggests that meanings could be identified with *intensions*, which are generalizations of extensions designed to cater for precisely the distinction that (5) brings out. We write $^\wedge e$ for the intension of an expression *e*. (5) shows that 'cordate' and 'renate' are only *contingently* coextensive: their extensions coincide in the actual world, but at other possible worlds, for instance ones where some cordates lack kidneys, their extensions diverge. The intension of a predicate, on the usual account, is its *possible-worlds profile*. This is not the set of things it applies to in the actual world, but rather a collection of pairs, each pair consisting in a possible world *w* and the set of things the predicate applies to in *w* (these are the things the predicate would have applied to if things had gone as they do in *w*).

Writing '@' for the actual world, the intension of a predicate *F* has the shape

(7) $\{\langle @, \{a_1, a_2, a_3, \ldots\}\rangle, \langle u, \{b_1, b_2, b_3, \ldots\}\rangle, \langle v, \{c_1, c_2, c_3, \ldots\}\rangle, \ldots\}$.

$\{a_1, a_2, a_3, \ldots\}$ is the (actual) extension of *F*. $\{b_1, b_2, b_3, \ldots\}$ is the set of things which would have been *F* if things had gone as they do in *u*, $\{c_1, c_2, c_3, \ldots\}$ is the set of the things which would have been *F* if things had gone as they do in *v*, and so on. In terms of our running example, we then have $^\wedge$cordate $\neq$ $^\wedge$renate, because for at least one *w*, there will be sets *X* and *Y* such that $X \neq Y$, $\langle w, X\rangle$ is in $^\wedge$cordate, and $\langle w, Y\rangle$ is in $^\wedge$renate. Since $^\wedge$cordate $\neq$ $^\wedge$renate and we wanted ⟦cordate⟧ $\neq$ ⟦renate⟧, we are encouraged to identify meaning with intension, at least for predicates.

Intensionalism about predicates solves the pet-fish problem as well as extensionalism does. To satisfy principle (i) on page 3, we can explain composition of predicates in terms of second-co-ordinate intersection for corresponding pairs. In other words, if $\langle w, X\rangle$ is in ⟦pet⟧, and $\langle w, Y\rangle$ is in ⟦fish⟧, we will have $\langle w, X \cap Y\rangle$ in ⟦pet⟧(⟦fish⟧). But $X \cap Y$ is exactly the extension of 'pet fish' at *w*. Therefore, we also have $\langle w, X \cap Y\rangle$ in ⟦pet fish⟧, and conversely. So principle (ii) on page 3 is also satisfied.

The predicate's intension is said to be *flexible* if its extension at some world is a different set from its extension at some other; all the predicates we have used so far, 'cordate', 'pet', 'fish' and 'renate', have flexible intensions. Kripke (1972, 1980) has convincingly argued that by contrast, the intensions of proper names are inflexible or *constant* (names are, in his terminology, *rigid designators*). So the intension of a name will have the form $\{\langle @, x\rangle, \langle u, x\rangle, \langle v, x\rangle, \ldots\}$, where x is a fixed object. This captures the fact that if we say

(8) Socrates could have been a cordate without being a renate

we are making a claim about Socrates, not someone else, so we are keeping the extension of 'Socrates' fixed.[6] But we must allow the extensions of 'cordate' and 'renate' to come apart at some world, otherwise (8) would be refuted. Granted that 'cordate' and 'renate' are actually co-extensive, at least one must be flexible if (8) is to be true.

The constancy or otherwise of the intensions of other types of expression will depend on the constituents of the expression. For sentences, necessary truths and necessary falsehoods have constant intensions, while contingencies have flexible ones. The intension of a sentence is a collection of pairs, each pair consisting in a possible world and the truth-value of the sentence at that world. So the intension of a necessary truth has the form $\{\langle @, \top\rangle, \langle u, \top\rangle, \langle v, \top\rangle, \ldots\}$. The intension of a sentence is sometimes called the *proposition* it expresses. The proposition, in this technical sense, can also be identified with a set of possible worlds, on the understanding that to say that a world w is in a proposition p is to say that $\langle w, \top\rangle$ is in p, and to say that w is not in p is to say that $\langle w, \bot\rangle$ is in p.

[6] To forestall any confusion: Kripke is of course happy to agree that there are ways things could have gone in which someone other than Socrates is *called* 'Socrates' (Kripke 1980:77–8). We can say, if we like, that there are worlds *in* which 'Socrates' does not denote Socrates: Socrates' parents choose a different name for him, and maybe his brother gets the name 'Socrates'. But (8) is true iff the indicative sentence 'Socrates is a cordate but not a renate' is made true by some possible world. In evaluating this sentence at a given possible world, we interpret 'Socrates' as standing for Socrates (who else?). But we need not interpret 'cordate' and 'renate' by their actual extensions. Since 'Socrates' is always interpreted by Socrates, we can say that it stands for Socrates *at*, or *with respect to*, every possible world. (For 'Socrates might not have existed' to be true, we need a world at which 'Socrates does not exist' is true. For uniform treatment with 'Socrates is a cordate but not a renate', we should still take 'Socrates' to denote Socrates with respect to such worlds.)

More generally, the intension $^\wedge e$ of any expression e is the set of pairs $\langle u, {}^{\vee u}e\rangle$, where ${}^{\vee u}e$ is the extension of e at the world u. For example, the intension of 'three blind mice' is the set of pairs $\langle u, X\rangle$, where X is the set of all sets in u that include three blind mice of u. Since 'blind mouse' has a flexible intension, so does 'three blind mice'. There are intensionalist theses for specific classes of expression in a language $\mathcal{L}$, and also global intensionalism, which says that for any meaningful expression e of $\mathcal{L}$, $[\![e]\!] = {}^\wedge e$.

Call an expression e_1 *extensional* iff for any expression e_2 such that $e_1{}^\frown e_2$ is meaningful, the extension of $e_1{}^\frown e_2$ depends on the *extension* of e_2, as opposed to its intension: ${}^\vee(e_1{}^\frown e_2) = {}^\vee e_1({}^\vee e_2)$. Call e_1 *intensional* iff the extension of $e_1{}^\frown e_2$ depends on the *intension* of e_2, that is, iff ${}^\vee(e_1{}^\frown e_2) = {}^\vee e_1({}^\wedge e_2)$. This is a distinction that is orthogonal to extensionalism/intensionalism: the meaning of every expression may be its extension, its intension, or something else, but expressions still *have* extensions, and the extension of a compound expression $e_1{}^\frown e_2$ may or may not be a function of the extensions of both e_1 and e_2.

'Pet' is extensional, since the extension of, say, 'pet fish', is obtained by intersecting the extensions of 'pet' and 'fish' – the *intension* of 'fish' does not get involved. But 'it is necessary that', as in

(9) It is necessary that all cordates are renates,

is intensional. To calculate the extension of (9), that is, its truth-value, we need the *extension* of 'it is necessary that' and the *intension* of 'all cordates are renates'. The extension of 'it is necessary that' is such that when 'it is necessary that' is combined with a sentence, the resulting expression (e.g., (9)) has a truth-value as its extension. But the relevant feature of the embedded sentence is its *in*tension. For example, the extension of 'all cordates are renates' does not suffice to fix the extension of (9), given the extension of 'it is necessary that': the extension of (9) depends on what type of intension 'all cordates are renates' has. If its intension is flexible, as we proposed in connection with (5a), then (9) is false, but if it is constantly true, making 'all cordates are renates' a necessary truth, then (9) is true. More explicitly, the extension of 'it is necessary that _' is such that when 'it is necessary that _' composes with a complete sentence S, the resulting complex sentence is true iff the intension of S is the set of pairs $\langle w, \top\rangle$ for every possible world w.

5 BEYOND INTENSIONALITY

Examples like (5) appear to show that extensions are inadequate candidates for meanings. Are there examples of a similar nature, which show that intensions are inadequate candidates? Apparently so, as the following pairs attest.

(10) a. Lex Luthor fears Superman.
 b. Lex Luthor fears Clark Kent.

(11) a. No-one doubts that water is water.
 b. No-one doubts that water is H_2O.

(12) a. Obviously, there are infinitely many numbers.
 b. Obviously, there are infinitely many prime numbers.

'Superman' and 'Clark Kent' have the same intension (assume the fiction to be factual), as do 'water' and 'H_2O',[7] and as do 'there are infinitely many numbers' and 'there are infinitely many prime numbers'. So if meaning is intension, we have ⟦Superman⟧ = ⟦Clark Kent⟧, ⟦water⟧ = ⟦H_2O⟧, and ⟦there are infinitely many numbers⟧ = ⟦there are infinitely many prime numbers⟧. Hence by compositionality, the meaning of each (a)-sentence above is the same as its companion (b)-sentence. So the members of each pair must have the same truth-values. But in each pair, the (a)-sentence appears to be true and the (b)-sentence false. Granted appearances, it follows that identifying meaning with intension is a mistake.

[7] 'H_2O' is a rigid designator because of how it is related to the rigidly designating description, 'the substance whose chemical composition is two parts of hydrogen to one of oxygen'. This description rigidly designates because (i) 'hydrogen' and 'oxygen' rigidly designate and (ii) when two parts of hydrogen combine with one part of oxygen in one possible world, the result is the same substance as in any other world where the same combination occurs. Since this substance is water in the actual world, it is water in every world where the combination occurs. However, to get the same intension for 'H_2O' as for 'water', granted the rigidity of the latter, we also require that water could not have had a different chemical composition, or more broadly, that the fundamental physical properties of substances (let's take chemical composition to be fundamental) are essential to them, and we need to be willing to treat the description 'H_2O' and the unstructured term 'water' the same way *vis à vis* worlds where there is no water, hydrogen, or oxygen.

However, the three examples are rather disparate. In (11) and (12), the differences in meaning are based on differences in complexity. The term 'water' is unstructured, and the term 'H_2O' is structured. The former can be mastered without any grasp of 'hydrogen', but not the latter (assuming it is not learned as an unstructured term). Similarly, the difference in meaning of the sentences in (12) traces to the presence of '*prime* number' in (12b). But for (10) there are no structural differences to cite in explaining why ⟦Superman⟧ ≠ ⟦Clark Kent⟧: each name is a semantic primitive. So it is natural to posit *hidden* differences. However, this leads to a dilemma. The more recondite we make the hidden difference, the more doubtful it is that it is grasped, even implicitly, by ordinary speakers. But a less recondite difference runs the risk of resembling the 'famous-deeds descriptions' theory discredited in (Kripke 1972, 1980).

In fact, the only candidate we have so far come up with for the meaning of a name is the object the name stands for.[8] But since Superman is Clark Kent, this would give those two names the same meaning, and promote the counterintuitive conclusion that the sentences in (10) cannot differ in truth-value. This is a problem to which we will return in Chapter 8.

6 HYPERINTENSIONAL LOGIC

An expression e is said to be *hyperintensional* iff there are expressions e_1 and e_2 of the same syntactic category such that $e^\frown e_1$ and $e^\frown e_2$ are meaningful, e_1 and e_2 have the same intension, but the intensions of $e^\frown e_1$ and $e^\frown e_2$ are different ($^\wedge e_1 = {}^\wedge e_2$ but $^\wedge(e^\frown e_1) \neq {}^\wedge(e^\frown e_2)$). We can use examples (10), (11), and (12) to show that their psychological vocabulary is hyperintensional, for instance, by putting 'fears' for e, 'Superman' for e_1, 'fears Superman' for $e^\frown e_1$, 'Clark Kent' for e_2, and 'fears Clark Kent' for $e^\frown e_2$. By compositionality for intensions,

[8] The intension of a name is not its referent α but rather, the set of pairs $\langle w, \alpha \rangle$ for each world w. However, this is simply a technical convenience, and should not distract attention from the fact that it is the referent that does the work in distinguishing one intension from another. It is trivial to convert an intensional theory in which names have intensions which are sets of pairs into one in which the intension of a name is identified with its extension.

^(fears Superman) ≠ ^(fears Clark Kent), otherwise composing each with ^(Lex Luthor) would produce the same intension. But granted that (10a) and (10b) have different truth-values at the actual world, they have different intensions. Thus 'fears' is a hyperintensional transitive verb by the criterion.

If we were to proceed as before, we would now discuss the hypothesis that the meaning of an expression is to be identified with its *hyper*intension. But there is no commonly accepted characterization of hyperintensions that leaves it open whether or not meaning is hyperintension. Nevertheless, we can use (11) and (12) to draw a conclusion about what hyperintensions must be like: they must not collapse evident differences in structure and constituents. In particular, the hyperintensions of 'water' and 'H_2O' must in some way reflect the fact that only the second of these involves ⟦hydrogen⟧, and the hyperintensions of 'there are infinitely many numbers' and 'there are infinitely many prime numbers' must differ in a way that traces to the presence of ⟦prime⟧ in the second but not the first. What to say about the cases in (10) is not settled at this point.

There are, broadly, two ways in which these distinctions might be realized. On one approach, the hyperintension of an expression is a structured entity that reflects the structure of the expression itself: the meanings of the primitive expressions are the basic constituents of the hyperintension of the complex expression, and they are configured in some set-theoretic arrangement that mirrors the syntax of the complex expression. A version of this says that the hyperintensions of names are their referents, and the hyperintensions of predicates are the properties they express. The hyperintension of a sentence is said to be a *state of affairs* or a *Russellian proposition*.[9] But on this account, (10a) and (10b) have the same hyperintension, which, some would say, defeats the hypothesis that hyperintension is meaning (we will see, in Chapter 8, why this is wrong).

A distinct, non-structuralist, approach, takes sentence-meanings

[9] See (Forbes 1989:137–149) on states of affairs, and (Soames 1987) on Russellian propositions. There is a hybrid account in (Lewis 1972:182–6) on which a sentence-meaning is a tree at each node of which we find the intension of the corresponding phrasal constituent of the sentence. So presumably there is also a hybrid account of meaning-composition, as tree-composition and intension-composition.

to be primitive (though there is no harm in thinking of them, extra-theoretically, as structured). In part, the motivation is that taking the notion of truth-value as primitive does not lead to an adequate conception of sentence-meaning, and taking the notion of possible world as primitive effects insufficient improvement. So it may be a mistake to try to identify sentence-meaning with something that is in some sense "relatively less problematic". And while it might seem to rob a semantic theory of interest if it avoids any such identification, that is not so, for there is still the task of finding meanings for subsentential units which compose together to result in the meanings of the sentences that they form. In fact, we get a pure form of the idea that the meaning of a subsentential unit is its contribution to the determination of the meanings of the complete sentences in which it occurs. And there are plenty of cases, some the focus of this book, in which the contribution is obscure.

One version of the non-structuralist approach is algebraic in nature (Bealer 1989, 1994). But there is another version, whose appeal rests on its very lucid *function-argument* model of meaning-composition. In Section 1, we wrote $x(y)$ for the composition of meanings x and y. This notation is silent about what composition consists in, and is consistent with it just being juxtaposition. But juxtaposing two meanings does not guarantee that they will compose together: the meanings have to be, in some sense, "suited" to each other. The function-argument model supports a clear account of what suitedness consists in for x and y: one meaning is a function that allows certain inputs ('arguments') and disallows certain others, and the other meaning is one of the allowed inputs. In the notation $x(y)$, x is the function and y is the argument or input.

This model can be applied to extensions, where it gives rise to what is known as the simple theory of types; it can be applied to intensions, producing higher-order intensional logic; and it can be applied to hyperintensions, where it produces what in (Thomason 1980) is dubbed 'intentional logic', and will be called 'hyperintensional logic' here (Thomason's system distinguishes (10a) and (10b), which our variant does not; see further n. 12, page 159). Our focus is on hyperintensional logic, but rather than leap in at the deep end, we start with a review of the function-argument model for extensions; readers familiar with the ideas can skip Section 1 of the next chapter.