1.2.1 Top-Down Metaphysics

Quine’s method for metaphysics and his “criterion of ontological commitment” are part of the background against which contemporary metaphysics is conducted. But, perhaps for this reason, the method is often left unstated, and so understood and either adopted or resisted in different versions. Problems about its formulation lead P. van Inwagen, in “Fictional Entities” to say that “there is no proposition, no thesis, that can be called ‘Quine’s criterion of ontological commitment’” (143). He thinks that, insofar as anything deserves the name ‘Quine’s criterion of ontological commitment’ it is a strategy or technique, not a thesis, and embraces it as such. (But it is not clear how a strategy should be less amenable to clear formulation than a thesis.) Nevertheless, I offer an account of the method and apply it in a pair of cases with platonistic results. These cases of properties and fictional objects are developed from van Inwagen, and are ones to which we shall return. I react to van Inwagen, not because I think his reasoning is flawed — quite the opposite! Van Inwagen’s development is particularly nice for our purposes insofar as he is clear-headed both about positive results, and limitations.

**Quine’s Method.** Though Quine himself may not have been content to start this way, let us suppose there is a world and that ordinary claims and theories are true and false by virtue of the way it is. Then we get at what there is by saying what in the world makes our claims and theories true — since, if the claims and theories are true, something in the world makes them true. Faced with some true claim or theory, the conditions under which it is true may be less than obvious. But we may offer an account of that in the world which makes it true. So the ordinary claims and theories are data with metaphysical results. Quine imagines that we offer accounts of truth conditions in an extensional language along the lines of that from Frege and Russell. Given this, his approach to ontological commitments is characterized by a pair of key theses.

Quine is notorious for having stated multiple, not entirely equivalent, formulations of his first thesis. However, for our purposes, it is clear enough what he has in mind. In “On What There Is” he says, in slogan form, that “to be is to be the value of a variable” (15) and more explicitly,

Q1 A theory is committed to those and only those entities to which the bound variables of the theory must be capable of referring in order that the affirmations made in the theory be true (13-14).
The thesis has both a positive and a negative side. In “On What There Is” Quine begins by emphasizing the negative side. He thinks he is not automatically committed to objects of reference by the use of proper names, and not committed to things like meanings or properties by the use of predicates. We need not pause here, as there are problems enough from the positive part. We are positively committed to whatever things must be in the range of variables for our affirmations to be true. Thus, to take an example from Quine, if the truth condition for “some zoological species are cross-fertile” is, $\exists x \exists y (x \text{ is a zoological species } \land y \text{ is a zoological species } \land x \neq y \land x \text{ is cross-fertile with } y)$ then we are committed to zoological species as such. But if we are able to offer some other account which does not require species in the range of the variables, say, $\exists x \exists y (x \text{ is an animal } \land y \text{ is an animal } \land x \neq y \land x \text{ is not the same species as } y \land x \text{ is cross-fertile with } y)$ then so long as the account is adequate in other respects, we might be committed just to the animals required for its truth. Quine speaks of “paraphrasing the statement as to show that the seeming reference to species on the part of our bound variable was an avoidable manner of speaking” (13). But, whatever “paraphrase” may come to, we have at least competing theories of the conditions under which the original claim is true, with divergent commitments. Observe that, so long as the truth conditions of expressions are preserved, and an argument is valid just in case there is no condition under which premises are true and the conclusions are not, consequences of expressions should be preserved as well — for validity hinges on nothing more than truth conditions. So a test of adequacy for this project of “regimenting” truth conditions is that consequences of expressions are retained.

We thus employ Quine’s method when we move from data including theories or claims regarded as true, to accounts of the conditions under which they are true, and from such accounts to ontological commitments. I have suggested that Quine supposes accounts of truth are given in something like the extensional (canonical) notation of Frege and Russell. But this suggests a second thesis.

Q2 The truth condition for any expression (which has a truth value) may be expressed in some extensional language.

Suppose this is false, that there are true expressions whose truth condition has no expression in an extensional language. Then the situation is as follows,
where (A) includes all expressions with a truth condition and (E) expressions with an
extensional truth condition. (E) is a subset of (A). But then Q1 itself would seem to
be problematic, insofar as there is room for true theories whose commitments are not
those to which bound variables of an extensional language must be assigned. Faced
with some recalcitrant or problematic bit of data, one might “opt out” of the method,
on the ground that data falls into the range to which the method does not apply. But
Q2 closes this gap. Given this, in Russell’s apt phrase, opting out might seem to have
“many advantages; [but] they are the same as the advantages of theft over honest toil.”

Q2 is particularly significant when data seems non-extensional. Extensional
expressions are characterized by stability of truth value under certain substitutions.
Deriving notions of singular terms, predicates and sentences from the canonical
notation, let us say singular terms are co-referential when they pick out the same
object, predicates are co-referential when they apply to the same objects, and sentences
are co-referential when they have the same truth value. Then,

EX A sentence is extensional iff switching a singular term, predicate, or sentential
part for one with the same reference cannot alter the truth value of the whole
sentence.

For a stock example, suppose ‘is a creature with a heart’ and ‘is a creature with a
kidney’ apply to the same individuals — where Lois believes that Superman has a
heart, but being unaware of Kryptonite biology, does not believe that he has a kidney.
And compare, ‘Lois believes Superman is a creature with a kidney’ with ‘It is not the
case that Superman is a creature with a kidney’. Under our assumptions, both are
false. But switching the co-referential ‘Superman’ with ‘Clark’ flips the first to true,
and leaves the second unchanged; switching the co-referential ‘is a creature with a
heart’ for ‘is a creature with a kidney’ again changes the first to true, and leaves the
second unchanged; and either switch has the effect of changing the sentential part,
‘Superman is a creature with a kidney’ for one with the same truth value. So the first
fails each of the conditions for extensionality, and the other none. Expressions from
the canonical notation all satisfy EX. But other languages might do so as well. Certain

9From a related context in “The Philosophy of Logical Atomism,” (71). In a nice example of
such theft, van Inwagen takes Ernest Gellner to task for saying he is a nominalist but does not try to
eliminate quantification over abstract objects from his discourse; “I do not try to put what I say into
canonical notation, and do not care what the notation looks like if someone else does it for me, and do
not feel in the very least bound by whatever ontic commitments such a translation may disclose” (“The
Last Pragmatist,” 203). Against this, van Inwagen restricts himself to the observation that, “Gellner’s
confession comes down to this: I don’t mind contradicting myself — I don’t mind both saying things
that imply that there are abstractions . . . and saying that there are no abstractions — if figuring out how
to avoid contradicting myself would require intellectual effort” (“Fictional Entities,” 144).
fragments of ordinary language are extensional, as might be expressions of languages with plural quantifiers, or infinitely long expressions. (Both plural quantification and infinitary languages are taken up in the following. Presentations for infinitary logic get technical quickly; however Nadel, “L₁ and Admissible Fragments” and Dickmann, “Larger Infinitary Languages” are a reasonable gateway. Oliver and Smiley, Plural Logic develops the other.)

Quine might justify Q2 as a theoretical or experimental result: as a matter of fact, when we set out to give truth conditions for the sentences of science or whatever, we find that expressions of the canonical notion, or of some (maybe extended) extensional language suffice. On the face of it, this suggestion is implausible — or, at least, Quine uses the fact that supposed truth conditions for certain expressions have no adequate extensional account as reason for questioning the intelligibility of those sentences. Thus in his classic, “Reference and Modality” he reasons from problems of substitutivity in modal contexts to his conclusion that results are “so much the worse for quantified modal logic” and “for unquantified modal logic as well” (156). The reasoning is hardly uncontroversial! My point here is just that data from modal logic, together with the negative results for extensionality, would seem to be just the sort of evidence that should call an experimental justification for Q2 into question.

But there may be other reasons to accept Q2. Quine begins “Reference and Modality” linking sameness of truth value on substitution of singular terms to the indiscernibility of identicals. By the indiscernibility of identicals, if \( a = b \), the properties of \( a \) are the same as the properties of \( b \). So, given a picture of language as picking out things and saying that they have whatever features they do, from \( a = b \) and \( Fa \) we expect \( Fb \). We have assumed that there is a world and that sentences are true or false by virtue of the way it is. Without prejudicing the question of what sorts of things there are, and of the nature of their properties or relations, one might think that this assumption amounts to saying that there are things with properties and relations, and that it is the things with their properties and relations that make sentences true or false. Say this is right. Then all we need to describe the world is the ability to pick out things and to say that they have whatever properties and relations they do. It may be that the truth condition for some sentence is that things or the world satisfy some complex condition; but so long as we are able to state this condition, there is nothing more to be said about the world. This seems to be just the sort of thing extensional languages are fitted to do. And it motivates Q2.

Now consider what happens when extensionality seems to fail. “Reference and Modality” opens with the point about the indiscernibility of identicals, and then an example where substitutivity seems problematic. Based on the true identity, ‘Cicero = Tully’, we might substitute into,
‘Cicero’ contains six letters
to obtain, “‘Tully’ contains six letters.” But the former is true and the latter is false. However this is not held to be a failure of substitutivity, insofar as nothing in the displayed sentence names Cicero. Rather, what is named is a word that contains six letters. A proper substitution is based on “‘Cicero’ = Tully’s other name that begins with ‘C’”, to reach the perfectly true, “Tully’s other name that begins with ‘C’ contains six letters.” By requiring that our language be extensional, we thus require clarity about what things are said to have what properties. And this is just right when we are concerned about questions of ontology.

More substantively, suppose we let a language have an operator for \( F \) for “at some future time” and that as it happens, ‘is Bob’s brother’ and ‘is Bob’s sibling’ are co-referential, although Bob has a sister on the way. In this case, \( F \exists x (x \text{ is Bob’s sibling } \land x \text{ is female}) \) is true though \( F \exists x (x \text{ is Bob’s brother } \land x \text{ is female}) \) is false. So truth value flips upon substitution of co-referential terms and extensionality fails. A response, like the one above, is that we have not gotten the objects and properties right. Thus, \( \exists t \exists x (t \text{ is a time after now } \land x \text{ is Bob’s sibling at } t \land x \text{ is female at } t) \) is true, while \( \exists t \exists x (t \text{ is a time after now } \land x \text{ is Bob’s brother at } t \land x \text{ is female at } t) \) is false. But there is no failure of extensionality, insofar as there is a sister and time to which ‘x is Bob’s sibling at t’ applies but ‘x is Bob’s brother at t’ does not; so the relations are not co-referential, even though they apply to the same persons now. Of course, among the objects over which the quantifiers range are times.

Little changes if the we consider an operator \( \diamond \) for possibility where \( \diamond \exists x (x \text{ is Bob’s sibling } \land x \text{ is female}) \) is true, though \( \diamond \exists x (x \text{ is Bob’s brother } \land x \text{ is female}) \) is false. Where ‘x is Bob’s sibling’ and ‘x is Bob’s brother’ actually apply to the same individuals, extensionality fails. One response is to let the quantifiers range over worlds. Thus, \( \exists w \exists x (w \text{ is a way the world could be } \land x \text{ is Bob’s sibling at } w \land x \text{ is female at } w) \) is true, while \( \exists w \exists x (w \text{ is a way the world could be } \land x \text{ is Bob’s brother at } w \land x \text{ is female at } w) \) is false. But there is no failure of extensionality, insofar as there is a sister and world to which ‘x is Bob’s sibling at w’ applies but ‘x is Bob’s brother at w’ does not; so the relations are not co-extensional, even though they overlap in actuality. And we quantify over ways the world can be.

Of course these are not the only responses. The appeal to other worlds is particularly problematic, and we shall be interested in alternatives in what follows. Quine

---

In a dismal episode from graduate school, I was trying to reproduce this example on a qualifying examination. Unfortunately, and typically, I mispelled Cicero’s other name ‘Tulley’. But then it did contain six letters! Flummoxed, I ended up creating an entirely separate example to make the point. Despite this sort of thing, I somehow managed to pass.
is satisfied enough with the response for times, and with corresponding ontological commitments. But he is not at all happy with the appeal to worlds. Our point for now is simply to emphasize the way Q1 combines with Q2 to pressure the way we think things are. Within the method, one is pressed to (i) accept the commitments, (ii) reject the data, or (iii) offer an alternative account of the truth conditions. In the ordinary case original data is secure — more secure than philosophical theories proposed to account for it. So we are engaged in the project of offering theories to account for truth, and so to identify corresponding ontological commitments.

Fictional Objects. In a series of articles spread over some 25 years, van Inwagen applies Quinean criteria for the conclusion that some things are fictional characters. Van Inwagen does not think Tolkien, for example, says something true or false when he says, in the first line of The Lord of the Rings, “When Mr. Bilbo Baggins of Bag End announced that he would shortly be celebrating his eleventy-first birthday . . . there was much talk and excitement in Hobbiton.” It would not make sense for someone peering over Tolkien’s shoulder when he was writing this to say, “How true!” or, “No, no, you’ve got it all wrong, that’s false!” (cf. “Creatures of Fiction,” 41). However expressions of the sort,

There are characters in some nineteenth-century novels who are presented with a greater wealth of physical detail than is any character in any eighteenth-century novel.

have truth values. That is, expressions of what we might call, “literary criticism” about fiction have truth values. As such, they are objects to which Quine’s method applies. So, a natural proposal is that the above expression has a structure of the sort, $\exists x(x \text{ is a character in a nineteenth-century novel} \land \forall y(y \text{ is a character in an eighteenth-century novel} \rightarrow x \text{ is presented with a greater wealth of physical detail than is } y)$. Among the things that must be assigned to the variables for this to be true are some characters in nineteenth-century novels. So by Q1 we are so-committed — or we are so-committed unless we want to reject the original data (and all other data of the sort), or can provide an alternative account which does not require the existence of the characters.

Van Inwagen is not content to reject the data, and does not think plausible alternative accounts of truth are forthcoming. So he accepts the commitment. It is difficult to demonstrate that no adequate alternative is possible. However van Inwagen emphasizes the constraint on such accounts mentioned above (“Creatures of Fiction,” 45f).
One might think that a commitment-free alternative is readily available. Seemingly, it is *novels*, or classes of novels, that do the work; so why not introduce a special relation, say, ‘*x dwelphs y*’ which applies iff members of *x* include a certain sort of “character development” more detailed than any in members of *y*. Then we say, as an account of the above, ‘the class of nineteenth-century novels dwelphs the class of eighteenth-century novels’ — and restrict our commitment to classes of novels. Presumably we get the truth values right. But van Inwagen asks us also to consider also,

> Every female character in any eighteenth-century novel is such that there is some character in some nineteenth-century novel who is presented with a greater wealth of physical detail than she is.

This is an obvious consequence of the original. And it follows also from the account of truth which quantifies over fictional characters. We might attempt an account that avoids characters as before — perhaps, “the class of nineteenth-century novels praphs the class of eighteenth-century novels.” But there is no obvious reason to think this follows from the claim that the class of nineteenth-century novels dwelphs the class of eighteenth-century novels. So the proposed alternative does not preserve logical consequence, and so far fails a condition of adequacy. Van Inwagen thinks accounts which preserve consequences “almost certainly . . . will have a quantificational structure not much simpler than the (apparent) quantificational structure of its ‘original’. ” Thus it is likely that commitments will remain.

Supposing that commitments remain, and so that there are creatures of fiction, there is the question what they are like. Presumably no physical things are hobbits, and no physical place is Middle Earth. But van Inwagen distinguishes properties fictional characters *have* from ones they *hold* (*in a text or location*). Fictional characters have logical properties like **BEING SELF-IDENTICAL**, along with literary properties like **BEING CREATED BY TOLKIEN** or **BEING A CHARACTER OF BOTH PRINT AND FILM**. They hold properties of the sort, **BEING SHORT**, **HAVING HAIRY FEET**, and **LIVING IN MIDDLE EARTH**. Thus it is van Inwagen’s view that “what appears to be the apparatus of predication in ‘fictional discourse’ is ambiguous. Sometimes it expresses actual predication, and sometimes an entirely different relation” (“Fictional Entities,” 148). When we say, “Frodo was created by Tolkien” we say that a certain property is possessed by the fictional character. When we say “Frodo is a hobbit” we say not that the character has, but that it stands in the holding relation to a property. So there is no pressure find short physical things with hairy feet to account for the truth of simple claims of the
sort, ‘Frodo is a hobbit’.\footnote{And similarly, ‘Frodo does not exist’ might be a way of saying truly that nothing has all the physical properties that Frodo holds. Or, for someone truly lost in the story, that ‘Frodo’ is not the name of a physical thing, but rather is a name held by a certain fictional character (“Fictional Entities,” 146-7).}

As van Inwagen observes, other theories may have the same general structure of allowing that there are fictional characters, with a distinction between properties the characters have, and ones they “hold.” He candidly agrees that he does not have much to say about the nature of characters such that they have some properties and hold others — and brings this out by comparison to the parallel but more specific theories of Wolterstorff and Thomasson. On Wolterstorff’s view, fictional characters are certain \textit{kinds}, where kinds are platonic entities (\textit{Worlds and Works of Art}). Very roughly, say one kind \textit{entails} another iff it is not possible for something to be of the first but not the second; then where properties are parallel to kinds, a character holds a property just in case it entails the property. This view has the consequence that fictional characters are eternal beings (or no less eternal than kinds), with all their properties essentially, discovered rather than created by their authors. In contrast, Thomasson holds that fictional characters are contingent beings, ‘abstract artifacts’ created by their authors (\textit{Fiction in Metaphysics}). A distinction like having and holding remains, but the nature of the objects is quite the opposite from the ones posited by Wolterstorff. Van Inwagen observes that these theories are subject to various concerns which he avoids “by the clever expedient of being vague” (“Fictional Entities,” 152). We shall return to this point shortly, when we turn to consideration of \textit{Problems About Primitives}.

\textbf{Properties.} Reasoning is parallel in van Inwagen’s, “A Theory of Properties.” So Quine’s method, backed by entailment considerations, has the result that there are properties. But the account of what they are turns out very thin. Van Inwagen takes as his example,

\begin{quote}
Spiders share some of the anatomical features of insects
\end{quote}

This has the apparent form, $\exists x (x$ is an anatomical feature $\land$ insects have $x \land$ spiders have $x$), which is true only if something in the range of the variables is an anatomical feature. But features, qualities, characteristics, properties, and the like may seem to be just the same thing. (And if there are distinctions to be made between any of these, it is likely that one will be no more palatable to the nominalist than another.) So by Q1 there is an apparent commitment to properties — unless, of course, we are willing to reject the data, or there is an acceptable way to account for the data that avoids the consequence.
In this case, the data seems secure — there is no denying that spiders share some of the anatomical features of insects. Again, it is impossible to survey all the attempts to account for the data. But the difficulty of providing alternatives is highlighted by the constraint that accounts of truth conditions preserve logical consequence. Consider a (reasonably traditional) response along the following lines: spiders share some of the anatomical features of insects just in case spiders are like insects in some anatomically relevant ways. We require a resemblance between spiders and insects. But van Inwagen observes that this seems to require a quantification over “ways one thing can be like another,” something like, $\exists x (x$ is a way one thing can be like another $\land x$ is anatomically relevant $\land$ spiders are like insects in $x$); thus there is commitment to ways one thing can be like another. And this may seem to be a perversion of the resemblance strategy. Perhaps the idea is merely to observe that there is an unstructured relation between the class of spiders and the class of insects so that, say, the class of spiders blaphs the class of insects. So far, so good. But consider the argument,

1. If two female spiders are of the same species, then one is like the other in all anatomically relevant ways.

2. If $a$ is like $b$ in some anatomically relevant way, and $b$ is like $c$ in the same way, then $a$ is like $c$ in that way.

3. An insect that is like a female spider in some anatomically relevant ways is like any female spider of the same species in some anatomically relevant ways.

The argument is valid. And if the premise and conclusion are given a structured account, then the conclusion follows by the usual methods. (Challenge: try it!) But it is hardly clear how the conclusion results where the premise and conclusion are unstructured. The point here is exactly parallel to the one for fictional characters. In effect, there is a bulge in the carpet: chasing it around moves the quantification from one place to another. In the end, van Inwagen thinks, the most comfortable place will be to rest where we started, with quantification over properties, and so with commitment to properties.

When it comes to saying what these things are like, van Inwagen offers a theory which, he says, is “nearly vacuous” (131). If we set out to describe the intrinsic nature of a pen or the like, we will have a great many things to say — about the nature of the ink, the working of the ball, or whatever. But not so for abstract objects in general, and properties in particular. Van Inwagen does, however, lay out a certain role which
is at least inconsistent with some things others have had to say about properties. His idea is to identify the property role with the role, “thing that can be said of something.” Some things can be said — as that Chicago has a population over two million. But this cannot be said of anything. However other things can be said of a thing, as of Chicago, or New York, or (falsely) of South Bend, that it has a population over two million. He calls things that can be said, or said of things assertibles; assertibles that can be said of things are unsaturated; and he applies the usual logical operations to these. Thus “that it has a population over two million” and “that it once filed an income-tax return” are unsaturated assertibles, as is “that it either has a population over two million or has filed an income-tax return.” We seem to quantify over unsaturated assertibles very much as we do properties as in, “all the negative things you’ve said about Gore are perfectly true, but don’t you see that they’re equally applicable to Bush?” (132). And it seems natural to identify unsaturated assertibles with properties.

But, so far, the picture is subject to paradox. If there are unsaturated assertables, then things can be said of them. Among the things we can say of “that it is white” for example, is that it is not white — the assertible is not a thing of the sort that has a color. So “that it is white” cannot be truly be said of itself. And similarly for many things that can be said of something — thus “that it has a population over two million” does not have a population at all, and cannot be truly said of itself. So, seemingly, one of the things that can be said of a thing is “that it cannot be truly said of itself.” But now there is trouble: Suppose this thing, “that it cannot be truly said of itself” is such that it can be truly said of itself; then for this reason it cannot be truly be said of itself. So suppose it cannot be truly be said of itself; then for this reason it can be truly said of itself. And this is impossible. Van Inwagen chooses to deny that one of the things that can be said of things is “that it cannot be truly said of itself.” As he observes, the existence of such an assertible may seem self-evident. However, so does the existence of a set whose members are just those sets that are not members of themselves. And anyone who accepts that there are sets will have to accept some response to Russell’s paradox. (Is the set of all sets not members of themselves a member of itself or no?) There are a variety of solutions — none, perhaps ideal. But whatever the solution may be, “the friends of things that can be said of things can easily adapt any of the standard, workable ways of dealing with the paradox to the task of saying which open sentences must correspond to things that can be said about things” (134).

Supposing, then, a reasonable response to the problem of paradox, the theory is left with some interesting consequences. Properties are not parts or constituents of concrete objects. Physical atoms and the like may be parts of a thing, but things that can be said of them are not among their parts. For this reason, properties are not “somehow more basic ontologically than the objects whose properties they are” (135).
If properties are things that can be said of something, then among the properties are ones that are complex like *singing in Vienna* or *made of ice* (Michael Jubien’s favorite property), so that properties are “abundant” rather than “sparse.” And among the properties are ones that are uninstantiated like *being round* or *not round*. Given these consequences (and more), van Inwagen remains concerned about the lack of content about their intrinsic nature. “The fact that this theory is inconsistent with various interesting and important theses about properties shows that, although it may be very close to being vacuous, it does not manage to be entirely vacuous” (138). Again, we shall return to this point, when we turn to consideration of Problems About Primitives.

Insofar as the Quinean motivation yields the existential claims about fictional objects, properties and the like apart from content about their intrinsic nature, it may seem so far neutral between hard and soft platonism. To the extent that nominalist theories of either sort fail to account for data about fictional characters, properties or the like, the platonist is likely to hold that nominalism simply fails. But, from a different starting point, the nominalist is likely to think the same about platonism. As we shall see, from the perspective of another methodological starting point, what counts as a compulsory question, and so the proper location of the burden of proof, may very well differ.\(^{13}\)

---

\(^{13}\)This point about burden of proof is a theme of the first section from Burgess and Rosen, *A Subject With No Object*. 