8

Preliminary Scouting Reports from the Outer Limits of Conceptual Engineering

Josh Dever

1. How to Argue about the Planets

Brown and Black are arguing about Pluto. Brown holds that Pluto is not a planet, in light of its similarities to paradigm non-planets such as Makemake, Eric, Haumea, and Quaoar. Black, on the other hand, holds that Pluto is a planet, in light of its similarity to paradigm planets such as Mercury, Venus, Earth, and Mars. (This is a tendentious and problematic way of describing the content of their disagreement. Be patient—it's early days, and Brown and Black will increase in philosophical sophistication soon.)

For a time, Brown and Black's dispute just amounts to citing various Plutonian features. Brown points out that the center of mass of the Pluto-Charon rotational system lies outside of Pluto. Black points out that Pluto has assumed a spherical shape due to its own gravitational forces. After a while, though, and after each discovering that the other is sometimes strangely unmoved by the considerations the first has advanced, they start to wonder what it is that they are arguing about. One point on which they quickly reach consensus is that they are *not* arguing about whether Pluto is characterized by the English word "planet". They both recall that a *Star Trek: Enterprise* episode (not surprisingly, a rather bad one) featured a rogue planet Dakala, and that Iain Banks' novel *Matter* is set on the artificially constructed shell planet Sursamen. They conclude that while the English word "planet" can apply to bodies that aren't orbiting a star and to human-made objects, nevertheless that *whatever* it is that they are trying to settle about Pluto, it's not whether *that* word applies to it, since they're both interested in some question that excludes the possibility of being an artificial or rogue object. (Brown and Black clearly still haven't gotten over their philosophical naivete and are moving rather too quickly from usage facts to meaning facts. But this won't matter much for we're heading. The point is that even if they had the right sort of evidence about the meaning of the English word "planet", they would in the face of that evidence still not be particularly interested in settling a question delimited by that meaning.)

Once they set aside the English word "planet", progress (of a sort) follows quickly. Brown suggests categories of PRIMARY PLANET and DWARF PLANET, with (e.g.) Earth and Mars in the first category and Pluto and Eris in the second. Black suggests categories of SOLITARY PLANET and BELT PLANET, with Earth and Pluto in the first category and Eris and Makemake in the second. At this point Brown and Black become tempted to describe their original argument as a "verbal dispute". Whether that's the right thing to say won't be a central issue here (although it's definitely running along tracks proximate to the path I want to explore), but it is right that Brown and Black agree that Pluto is a dwarf planet and a solitary planet and is not a primary planet or a belt planet.

Brown and Black's argument at this point might turn into an argument about *which concepts to use*. Brown suggests using PRIMARY PLANET and DWARF PLANET, and perhaps reserving the label "planet" for PRIMARY PLANET, and hence rejecting the sentence "Pluto is a planet". Black suggests using SOLITARY PLANET and BELT PLANET, and perhaps reserving the label "planet" for SOLITARY PLANET, and hence endorsing the sentence "Pluto is a planet". There can of course be many reasons to have such an argument, and many reasonable ways for such an argument to proceed. We are finite beings, so we have to pick and choose which theoretical projects to engage in, and we can look for reasons for suspecting that one of the two classificatory schemes proposed by Brown and Black will do a better job of furthering various of the goals that an astronomical theory is pursuing.

But there is also a clear sense of "ought" in which what we ought to be doing in this situation is just assembling all of the truths about PRIMARY PLANETS, DWARF PLANETS, SOLITARY PLANETS, and BELT PLANETS. No need for Brown and Black to disagree any longer. Brown can assemble his total theory of PRIMARY PLANETS and DWARF PLANETS, and Black can assemble his total theory of SOLITARY PLANETS and BELT PLANETS, and then we can dump these two theories together into a single megatheory. Brown, after all, agrees with everything Black has to say about SOLITARY PLANETS and BELT PLANETS, and Black agrees with everything Brown has to say about PRIMARY PLANETS and DWARF PLANETS.

2. Big Theory, Little Theory

We can characterize this point in Brown and Black's discussion in terms of the norms of theory selection. Let T1 be the total theory of PRIMARY PLANETS and DWARF PLANETS and T2 be the total theory of SOLITARY PLANETS and BELT PLANETS. Then T3 is the (logical closure of) the union of T1 and T2. How should theorists choose among T1, T2, and T3? There are two basic orientations toward the project of theorizing that we can distinguish here. We can be 'Big Theory' theorists, who think that the goal of theorizing is to say everything there is to be said. The Big Theory theorist favors T3 over T1 and T2. Why leave out the insights of either T1 or T2 when we could just say it all? Or we can be 'Little Theory' theorists, who think that the goal of theorizing is to say everything *worth saying*. The Little Theory theorist might, for example, favor T1 over T2 and T3 on the grounds that the additional information included in T2 and T3 isn't information

worth having (perhaps even is information that gets in the way of understanding the world).

The Big Theory theorist doesn't disagree with the Little Theory theorist that some information matters more than other information. But the Big Theory theorist has a 'just more theory' attitude toward that insight. So the Big Theory Theorist might want to add to T3 the further claim:

> PRIMARY PLANET and DWARF PLANET are more helpful categories for understanding astronomical matters than are SOLITARY PLANET and BELT PLANET.

If the Little Theory theorist is *right* in their practical preference for the T1 information over the T2 information, then that claim is true, and the Big Theory theorist thus wants to endorse it as well. The Little Theory theorist worries, however, that the 'just more theory' approach doesn't help separate the important from the unimportant, because we still don't distinguish the important from the unimportant auxiliary claims about rankings among primary claims. (One version of this worry is that if we are Big Theory theorists, we might include auxiliary claims about which of our primary claims are in a maximally natural vocabulary. But we'll also include auxiliary claims about which of our primary claims are in a maximally schmatural vocabulary, for some grue-like variant of natural, and we're then in danger of not knowing whether it's naturalness or schmaturalness that makes for importance. A bit more on this line of concern in the final section of this chapter.)

Of course, no one is really a Big Theory theorist in the simplistic sense set out above. We don't want our theory genuinely to say *everything* there is to be said. As Putnam (1978) observes, if there's anything that's clear, it's that not absolutely everything is to be endorsed. A decent first stab at *why* not everything should be endorsed is that endorsing everything requires endorsing all the falsehoods as well as all the truths, and we don't want to endorse the falsehoods. So a standard more plausible way of being a Big Theory theorist is to endorse the following norm of theorizing:

- **Alethic Theory Selection**: The sole criterion for theorizing is that all and only truths should be endorsed. (Since everything in all of T1, T2, and T3 is true, all of these claims should be endorsed. Since T3 is the meet of T1 and T2, endorsing everything in all three theories amounts to endorsing T3.)

Contrasting with a norm of alethic theory selection is:

- **Pragmatic Theory Selection**: Theories are selected (at least) in part on the basis of pragmatic factors such as explanatory power, computational efficiency, productivity in guiding future research, promotion of social welfare, and so forth. We might thus favor T1 over T2 and T3 on the grounds that the PRIMARY/DWARF PLANET distinction fits well with the principal classificatory aims of astronomy.

This truth-anchored picture of Big Theory theorizing is enough to get us going, but we'll see later that it will be helpful to have a way of thinking about things that isn't tied specifically to truth. We'll thus consider another way of putting things.

3. Conceptual Maximalism and Global Versus Ideal Language Theorizing

One attraction of being a Big Theory theorist is that there is a kind of winnowing down that we don't need to engage in. When confronted with the conceptual diversity of PRIMARY PLANET, DWARF PLANET, SOLITARY PLANET, and BELT PLANET, we don't need to make a choice. (*Qua* theorist. *Qua* person actually doing some investigating, we may need to decide which investigations to do now. But those decisions can be haphazard, idiosyncratic, and unprincipled.) We can be **conceptual maximalists**, welcoming the use of *all* concepts in our theorizing. But if we are pragmatic theory selectors, we do winnow down, since not *everything* goes into the final theory. The winnowing down then can be *haphazard* or *stratified*. Haphazard winnowing doesn't break down nicely along conceptual lines—we might accept some but not all (true) DWARF PLANET claims (where those claims have various pragmatic virtues) and also some but not all (true) BELT PLANET claims (where those have pragmatic virtues).

Standard pictures of theory selection, however, tend to assume that the winnowing will be stratified—that we will *first* pick out a pragmatically privileged collection of conceptual resources, and then *second* build a theory consisting of all true claims using those resources. On the stratified picture, we decide (on pragmatic grounds) to theorize in terms of DWARF PLANET rather than BELT PLANET, and having so decided, we then build the *full* theory of DWARF PLANET, containing all the DWARF PLANET truths.

Stratified pragmatic theory selection is then a form of **ideal language theorizing**. In ideal language theorizing, there is some privileged language, such that a theory couched in that language is theoretically preferred over a theory couched in another language. The ideal language might be the ordinary language of the theorizer (as in Hirsch (2002)), or a logically perspicuous fragment/regimentation of the ordinary language (as in Quine (1948)), or a non-ordinary language well-suited for metaphysical theorizing (Sider (2012)'s 'Ontologese' or a Lewis (1983)'s appeal to a language of maximally natural concepts).

We'll set aside haphazard pragmatic theory selection and focus on the dispute between alethic theory selection and stratified pragmatic theory selection. If stratified pragmatic theory selection is ideal language theorizing, alethic theory selection is **global language theorizing**. No one language is privileged; our theoretical obligation is to work in all languages (or in the meet of all languages, if there is such a language).

(One version of) Carnap is the distinctive global language theorist of our local tradition, providing a distinctive break with the various ideal language inclinations of Frege, Russell, Wittgenstein, and Quine. There are many Carnaps available in the interpretational space; I'm interested in the Universalist Carnap, who takes the external question to be without theoretical significance. What needs to be done is to answer all the internal questions in all the languages; the external question is just a decision about what portion of what needs to be done we are going to do now.

One advantage of recasting the Big Theory/Little Theory distinction as the global language/ideal language distinction rather than the alethic norm/pragmatic norm

distinction is that it lets us relax the specific role of truth in guiding Big Theory theorizing. We can take the target Big Theory to be the theory that lists all claims in all languages and for each claim specifies the semantic status of that claim. If 'true' and 'false' emerge as privileged semantic statuses across all languages, this will be isomorphic to the theory yielded by the alethic norm, but with the global language picture, we can engage in theorizing even if we don't know how to pick out a single status that across all languages marks claims as theoretically privileged. Our attention can turn from the normative questions raised by the alethic/pragmatic split to content-based questions raised by the global/ideal split. (A first look at one of those content-based questions will be our central concern here.)

(We shouldn't, though, be too sanguine about the non-alethic global language version of Big Theory theorizing. On this picture, a theory doesn't consist of a collection of claims, but rather of a collection of claim-status pairs. Given that the Big Theory response to Little Theory insistence that theories encode more than mere information (by, e.g., encoding what is important, fundamental, explanatory, and so on) was the 'just more theory' response that these other things could be included in the theory in the guise of further claims, a retreat from theory-as-claims is in danger of undermining a central motivation of Big Theory theorizing.)

4. Conceptual Engineering, Conceptual Ethics, and Xenolinguistics

I've couched things above in the language of some old disputes about metaontology. That was done in part to help highlight the (never very dimly illuminated) fact that the recent conceptual engineering movement is heavily indebted to those old disputes. To bring this portion of the two discussions fully into alignment, let's make the distinction between *conceptual engineering* and *conceptual ethics*:

- Conceptual engineering is the broadly theoretical task of designing new concepts (or changing our old concepts; I won't worry about the trans-engineering identity condition question here).
- Conceptual ethics is the broadly practical task of determining, post-engineering, which concepts we ought to use. (Note that conceptual ethics, as a species of pragmatic theory selection, requires that the pragmatism be of the stratified rather than the haphazard genus.)

The conceptual maximalist is an eliminativist about the project of conceptual ethics. There is no interesting question about which concepts we ought to use: we should use all of them, and build the maximal theory saying all that there is to be said in the most expansive language. That's a tempting eliminativism, especially for those of us inclined to think that the pragmatic considerations don't go deep enough to guide the kind of intellectual enterprise we take ourselves to be engaged in. (Again, this is eliminativism about conceptual ethics as a general issue in theory selection. That's an eliminativism that's compatible with serious and important questions about what *part* of the overall theoretical project any one person should engage in at any one time. But those questions, says the conceptual maximalist, belong in a

different disciplinary box. Our general answer to the question of how to theorize doesn't need to do anything to tell us whether Smith, in particular, should be a category theorist or a Marlowe scholar.)

But there is a price to be paid for the conceptual maximalist's eliminativism about conceptual ethics. If there is no ethical question here, and our theoretical task is to say what there is to be said using *all the concepts*, then we had better have something to say about what the range of possible concepts is. The way I've set things up here, that amounts to the question of what kinds of things count as *possible languages*. The central point I want to push here is that the question of what possible languages there are (equivalently, perhaps, what could be said in any possible language) is one of the deepest, most intractable, and most neglected philosophical issues out there. But, as I'll try to bring out, one of the things that makes the territory difficult here is that there's also a decent chance that the question is utterly trivial.

The conceptual maximalist/global language theorist (I'll assume henceforth that the two labels can be used interchangeably) dodges a specific conceptual engineering bullet that hits the non-maximalist directly: the conceptual maximalist doesn't need to decide which of the PRIMARY PLANET/DWARF PLANET and SOLITARY PLANET/BELT PLANET distinctions is more worthy of enshrining in our theory. But, of course, these aren't the only astronomical concepts available. There's also the PLANET WITH RINGS/PLANET WITHOUT RINGS distinction, and the STELLAR BODY WITH MORE SOLID THAN GASEOUS VOLUME/STELLAR BODY WITH MORE GASEOUS THAN SOLID VOLUME distinction, and many many more. Very quickly we reach the point of observing that a collection of N astronomical bodies immediately gives us 2^N astronomical distinctions. And even that's only the beginning. By adding temporal and modal intensional dimensions, our available distinctions presumably become robustly infinite, and numerous grue-ish categories emerge. The conceptual maximalist isn't *bothered* by the inclusion of grue-ish categories—there's no commitment to everything in the theory being projectable, and as usual the theory itself can include commentary on which tracked categories do project—but there is still a lot of theory to compile.

That sort of predicative plenitudinousness, though, only scratches the surface of possible languages. Minimally, we can make similarly plenitudinous moves at other semantic categories. There are quite a lot of generalized quantifiers, or adverbs, or modal operators, or $(((e,t),t),(((e,t),(e,t)),(t,e)))$-category expressions, that we can add to our expressive resources and then require the conceptual maximalist to theorize about. Such plenitudinous moves expand the language by adding more inputs to the same underlying semantic infrastructure—a basic term-predicate structure, or its generalization to a full categorical grammar. But a genuinely *global* language needs to entertain other more alien ways that languages could be structured. The theory of the global language theorist needs to include the contents of map-like representations and the contents of non-conceptual perceptual experiences. We need to work out whether there are claim-like (theory-worthy) contents captured using infinitary, non-well-founded, or gunky versions of semantic type theory, using scoreboard update procedures or strategies for scoreboard manipulation or equilibrium points for massively multi-player scoreboard scheming tactics, using relations between fundamental groups of non-Hausdorff topological spaces, using methods of manipulation patterns of social prestige markers, using functions from n-tuples of alien

phenomenal states to dispositions to adjust alien quasi-epistemic attitudes, and so on. Linguistics to date has been merely domestic linguistics; examination of the languages we happen to find around us. Global language theorizing calls for an ambitious xenolinguistics, in which we consider what languages there could have been. (In particular, that part of xenosemantics which is concerned not with the question of how xenomeanings are correlated with and derived from xenosyntax, but with what the xenomeanings are.) We haven't even started on the project of xenolinguistics.

Maybe we get lucky, and there's no real need for xenolinguistics. Maybe all of the potential exotica above, if they're capable of being deployed in the representational devices of a genuine language, end up encoding contents that are already made available by the semantic tools made available in English. If so, English is already a maximal language, capable of saying anything that can be said in any language. If not, there is at least local ineffability—claims that we can't express with English. There might be stronger versions of ineffability—claims that can't be expressed in any human language, or can't be thought or grasped by human minds. But whether there is ineffability or not, the conceptual maximalist, in order to say what would even count as the Big Theory, needs to answer:

- **The Boundary Question**: What is the range of, or the characteristic feature of, possible languages and possible things to be said in possible languages?

The Boundary Question isn't a question about what expressions could stand in the expressing relation to what contents, or a question about what it takes for speakers to be speakers of a language. (It's thus not asking about Lewis's *actual language* relation.) It's rather a question about what contents there are to be expressed. At a first draft, it can be taken as the question *what propositions are there?*, or *what does it take for a content to be a proposition?*, although I don't want to build in to the question the assumption that propositions are uniquely the kind of thing a language must express.

If you're a serious conceptual ethicist, you can avoid wrestling with the Boundary Question. If you go in for some conceptual ethics, your pragmatic norms can pick out a privileged ideal language of theorizing. With the language picked out, you've got your domain of theorizing in hand, and you won't need to explore the boundaries of conceptual possibility in order to carry out your theorizing task. In practice, what this comes down to is that your conceptual ethical concerns give you a prior picture of what kind of concepts might fulfill your theorizing goals, and so the conceptual engineering task gets constrained from the beginning. You might not know in advance what language will end up being the ideal one, but you can see *roughly* where the ideal language is located, and not venture too far outside that region in your engineering. DWARF PLANET and BELT PLANET get a look in, but even relatively parochial exotica such as IT PLANETIZES SOLITARILY and PERIHELION-PLANET/APHELION-PLANET (compare INCAR and OUTCAR) aren't the kinds of things we want our engineers to build.

Of course, not needing to answer the Boundary Question doesn't go make the Boundary Question go away. (If you're very lucky, your conceptual ethics might even make the question go away. Suppose your conceptual ethics are that we should adopt whatever concepts maximize overall human flourishing. (Black wins the planetary dispute, e.g., because he avoids saddening schoolchildren everywhere by depriving

Pluto of its status in the pantheon.) Perhaps then a language (a conceptual repertoire) just *is* a tool for maximizing overall human flourishing—in the end, when confronted with some practice and asked whether that practice amounts to a language, there is nothing more to do than to see whether that practice has as an aspect promoting human flourishing.)

And of course, there is some sense in which the conceptual maximalist doesn't have to answer the Boundary Question. There's lots of global language theorizing to do prior to engaging in some xenolinguistics. But an understanding of the full scope of the global language theorizing project, unlike an understanding of the full scope of the ideal language theorizing project, does call for an answer to the Boundary Question.

5. Problematic Languages and Limiting Damage and Exposure

The project for the remainder of this chapter, then, is to make some preliminary forays into answering the Boundary Question, mostly with an eye to demonstrating that easy answers aren't going to work. We'll consider various delimiting criteria stating what the range of possible languages is, and extract some overarching morals about the kinds of difficulties these criteria get into. To give a little extra punch to the Boundary Question, we begin by noting some potentially troubling commitments of conceptual maximalism, commitments that we will then hope that an adequate answer to the Boundary Question will help us avoid. The conceptual maximalist bears a theoretical commitment to build a maximal theory collecting up all of the truths using all of the concepts in all of the possible languages. When we are dealing with Brown and Black, the maximalist route of theorizing using *all* of DWARF PLANET, BELT PLANET, and so on looks pleasingly cosmopolitan. Even in more pragmatically loaded settings, the maximalist approach has an appeal. We might think that on due consideration the normatively weighty concept of RACISM is one of systematic ill-treatment based on perceived racial classification deriving from a history of institutional oppression and power inequities, making the idea of "reverse racism" incoherent. But we can also easily acknowledge the concept of RACISM* that drops the institutional oppression and power inequity requirements, and build a total theory that collects truths about both RACISM and RACISM*, while expecting that few *normative* truths will involve RACISM*. But things aren't always so easy for the conceptual maximalist. Consider the following:

- Will the conceptual maximalist be required to include all of the (insert your favorite racial/gender/religious/etc. slur here—I'll use "freethinker" both so that I'm in the slurred group and because I like reclaiming that eighteenth-century slurring feel to the term) truths in the total theory? We might have hoped that "Jones is a freethinker" and "Freethinkers lack a moral compass" are the sorts of things that the *bigot* is committed to, not the sort of thing that we as ideal theorizers are committed to. But if there is a FREETHINKER concept then the conceptual maximalist must use that concept in theorizing. (The conceptual maximalist can always hope that there are no FREETHINKER truths to include in

the total theory, despite the presence of the FREETHINKER concept. Probably "Freethinkers lack a moral compass" can be avoided. "Jones is a freethinker" is harder. And avoiding *any* FREETHINKER truths presumably calls for some logical revisionism.)

- Will the conceptual maximalist be required to include all of the SHERLOCK HOLMES truths in the total theory? The conceptual maximalist is going to mention a lot of strange and unfamiliar objects in the total theory—belt planets, incars, book-like objects that exist only while it is unethical to open them to page 37. But does the conceptual maximalist need to mention Sherlock Holmes in the total theory? We might have thought that "Sherlock Holmes is a detective" is the sort of thing *John Watson* is committed to, not the sort of thing we as ideal theorizer are committed to. But if there is a SHERLOCK HOLMES concept, then the conceptual maximalist must theorize using it. (As before, we can hope there are no SHERLOCK HOLMES truths to be included. As before, realizing that hope creates pressure for logical revision.)
- Will the conceptual maximalist be required to include all of the TONK truths in the total theory? Letting in *any* TONK truths looks dangerous, because given the constitutive inferential rules of TONK, once some TONK claims go in, *all* claims go in. We might have thought that TONK claims were the sort of thing the logical deviant was committed to, but not the sort of thing that we as ideal theorizers should countenance.
- Will the conceptual maximalist be required to include the Eiffel Tower in the total theory? Not (the familiar concept) THE EIFFEL TOWER, and not the physical Eiffel Tower as (say) a constituent of a Russellian singular proposition, but the tower itself as the 'propositional' content of an utterance in a possible language?

In general, there's a lot of potential weird conceptual junk out there, and it would be nice if the conceptual maximalist had a way to produce a respectable junk-free total theory. There are two approaches here. One approach is **Limit the Damage**. On this approach, we countenance the "defective" concepts, but we argue that those concepts, because of their defectiveness, don't manage to feature in any true claims, and thus don't get into the total theory. The other approach is **Limit the Exposure**. On this approach, we find grounds for declining to countenance the "defective" concepts. If there are no languages that use the (putative) concepts FREETHINKER, SHERLOCK HOLMES, and TONK, and no languages in which the Eiffel Tower is the content of a claim, then the conceptual maximalist has nothing to fear from these cases.

Clearly the plausibility of the "no such language" line is increasing as we proceed through the examples. After all, English (we might think) *does* contain the concepts FREETHINKER and SHERLOCK HOLMES, so those two cases at a minimum do represent parts of possible languages. But we shouldn't be too quick to conclude that English is indeed a language. Perhaps it merely *appears* to be a language, but is prevented from being one by its deployment of "defective concepts". (In the same way that we might say that *Sherlock Holmes is a detective* merely *appears* to be a thought.) In the end there's a probably a choice here between a more expansive use of the term "language", on which English definitely counts as a language and on which the conceptual maximalist is committed only to theorizing uses the resources of all languages

bearing some good-making feature, and a less expansive use of the term "language" which reserves the term as an honorific for cases in which the good-making feature is present. I doubt it matters which way we go (as befits a conceptual maximalist); I'll talk in the latter way henceforth.

I'm going to seek to **Limit the Exposure** rather than **Limit the Damage**. That's in part because I think the damage is hard to limit, and the Boundary Question needs to be answered even if we decide to limit damage (and answering the Boundary Question is a way to limit the exposure). It's also in part because **Limit the Damage** requires the specifically *alethic* formulation of the non-pragmatic theoretical enterprise, and (as hinted earlier, and as will come out soon) I think it's hard to hold on to the alethic formulation as we investigate the Boundary Question.

6. Answering the Boundary Question

So much for preliminaries. What, then, is the possible range of languages? We will consider two general strategies for answering the Boundary Question. One strategy is to extract an answer from our understanding of semantic theories—we check whether the existing semantic toolkit offers an answer to the question 'what sentence-level contents are available to be expressed" (in, e.g., the way that the standard semantic toolkit offers as an answer to the question 'what available quantifier-level contents are available to be expressed' the category $((e,t),t)$). The second strategy is to extract an answer from our *metasemantics*, by examining how our understanding of the theoretical role of contents delimits what entities could play that role.

First Semantic Attempt: Possible Worlds

Let's start with an answer that drops naturally out of one popular framework for semantic theorizing. Propositions, many people say, are sets of possible worlds. One nice thing about this doctrine is that it immediately tells us what the full range of propositions is: the power set of possible worlds. The maximal language, then, is the language that allows expression of each member of the power set. (Note that we're concerned with the language only with respect to what the language expresses, not with what vehicle it uses in doing the expressing. So even if, for Kaplan paradox reasons, we can't come up with enough vehicles to express all the contents, we needn't worry about that limitation for current purposes. We similarly won't be interested in the question of whether anyone could ever entertain all of the resulting contents, or even whether each of the contents is possibly entertainable, again sidestepping Kaplan paradox worries.) The conceptual maximalist is then committed to building a theory that settles each of these propositions (and which, as a result, decides exactly which world is the actual world). PLUTO IS A DWARF PLANET picks out one set of worlds containing the actual world (which thus goes in the total theory); PLUTO IS A SOLITARY PLANET picks out a different set of worlds containing the actual world.

I'll focus on two worries about this possible worlds answer to the Boundary Question—both selected because they are relatively easy versions of worries that will hound more sophisticated approaches as we go. The first concern as one of **Undue Expressive Limitation**. The possible worlds framework is of course notoriously coarse-grained in its expressive capacity. By giving the possible worlds answer

to the Boundary Question we are thereby saying, for example, that *no possible language* can have more than one logically true content, or can distinguish between Hesperus and Phosphorus contents. We rule out as *impossible* Fregean languages that have more than one concept of a given object or Russellian languages that distinguish among concepts based on their internal structure. Of course, fans of the possible worlds framework have things to say in response to these coarseness of grain considerations, and I don't mean them to be decisive. But I do think they are weightier when we consider whether possible worlds give us a suitable framework for *all possible languages* than when we simply consider how to analyze our own language.

The second and deeper concern is one of **Passing the Buck**. I'm assuming for now that our possible worlds framework is not a Lewisian "modal realist" framework (we'll touch on the Lewisian alternative below). As a result, we need some story about what possible worlds are. But all of the off-the-shelf stories seem to just raise the same Boundary Question problems again, simply slightly relocated. This is most obvious if we take possible worlds to be maximal modally compossible collections of propositions (but shows up also if we take possible worlds to be a partition of some basic/atomic/fundamental propositions, or to be maximal properties/states of affairs). Consider: is there a possible worlds content JONES IS A FREETHINKER? That depends on whether possible worlds themselves are characterized in part in FREETHINKER terms. We definitely want a possible world to settle who is an atheist and who is not. Do we also want a possible world to settle who is a freethinker and who is not? If we do, we'll find FREETHINKER propositions among our maximalist collection of possible worlds contents; if we don't, we won't. But the question of whether possible worlds settle who is a freethinker looks suspiciously like the question of whether there is a (real, non-defective) FREETHINKER concept or whether a total theory will settle who is a freethinker. Of course, paradox of analysis issues threaten here—these things had better be closely linked, given that we're trying to get an account of what the range of languages is that answers the Boundary Question and hence tells us what our ideal theory will discuss. But the territory has the definite odor of the non-explanatory regress and the unilluminating circle here. Similarly with other cases—do possible worlds settle TONK matters and SHERLOCK HOLMES matters? In picking out a possible world, do we need to specify whether (insert Eiffel Tower here)? In the end, I'm skeptical that we have any better grip on the notion of what a possible world is than on what a possible content is. Onward, then.

Second Semantic Attempt: Truth Conditions

The coarse-graining observed under **Undue Expressive Limitations** above suggests a natural next attempt: let's move to a more fine-grained semantic framework. There are many to choose from, but we'll consider the thought that a theory of meaning for a language takes the form of specifying truth conditions, so that contents can be equated with those truth conditions. This is an especially natural proposal in a setting in which we're tracking the prospects for a purely *alethic* criterion of theory selection. I'm not sure I know what truth conditions are (more on that momentarily), but many people seem to think that the truth conditions *true when 2 and 2 make 4* are different from the truth conditions *true when all groups of order 7 are abelian*, and even that

the truth conditions *true when Hesperus is visible* are different from the truth conditions *true when Phosphorus is visible*.

Nevertheless, truth conditional semantics can still be plausibly accused of **Undue Expressive Limitations**. There are many off-the-shelf semantic machineries available to witness the potential expressive limitations. To pick a few:

- Many semantic frameworks use *truth-like* features, such as truth at a world, at a time, at a point of assessment, or at some other index of evaluation; or a Tarskian hierarchy of truth predicates, or separate ascending truth and descending truth.
- Many semantic frameworks *supplement* truth conditions with other content features, as in Potts-style two-dimensional accounts that have a dimension of expressive meaning.
- Some semantic frameworks eschew truth entirely, as in Gibbard/Blackburn-style expressivism.

I don't mean any of these cases for expressive limitation to be decisive. I'm not presupposing a methodology of ecumenicism, on which our account of the range of possible languages needs to accommodate everything any theorist has proposed as a language—it may well be that people have been writing down machinery for things that don't in fact count as languages. But I do think they are at least indicative that the notion of truth isn't enough to give us everything there could be in a language.

More importantly, truth conditional semantics are still subject to **Passing the Buck** worries. We can see this preliminarily by asking whether the following are specifications of truth conditions:

- "Jones is a freethinker" is true iff Jones is a freethinker.
- "Sherlock Holmes is a detective" is true iff Sherlock Holmes is a detective.
- "Trump is president tonk Pence is vice-president" is true iff Trump is president tonk Pence is vice-president.
- (Insert Eiffel Tower here) is true iff (insert Eiffel Tower here).

Hopefully not, for at least some of these, but I don't see anything better to say about why not than the prior observation that some of the putative concepts involved are defective and aren't part of any real language. More generally, the question is whether we have a picture of what kind of property truth is that puts helpful limits on what kinds of things can be truth conditions. (Note that it's not the range of truth *bearers* that is at issue here, but the range of truth conditions born.) Perhaps we can agree that it's in the nature of truth that truth conditions have to be given via (contents of) T-sentences of the form:

- (Truth bearer) is true iff (truth condition)

But that's not helpful if we don't know what the range of (contents of) T-sentences is. On the one hand, we already have reason to think that English may be overly generous on this account. *"Sherlock Holmes is a detective" is true iff Sherlock Holmes is a detective* is a grammatical English sentence of the form of a T-sentence, but that can't be *decisive* on the question of whether SHERLOCK HOLMES claims have real truth conditions. On the other hand, it's also plausible that English is overly restrictive on

this account. If there are expressive limitations of English, of course, we should expect there to be truth conditions that can't be specified by English-language T-sentences. But it's also unclear what argument there is that the English syntactic category of sentence matches the range of permissible T-specifications (i.e., whether it's really distinctively T-*sentences* that we are out for).

Disquotational accounts of truth, for example, simply give no answer to the question of why (e.g.)

- "Aristotle" is true iff Aristotle.

Is not a valid instance of disquotation, beyond the simple insistence that the disquotational instances be sentences. (Is the reason that *"Aristotle" is true iff Aristotle* is uninterpretable (as opposed to ungrammatical))? I don't see why it would be, unless we have already decided that:

- "'Aristotle' is true iff Aristotle" is true iff "Aristotle" is true iff Aristotle.

is uninterpretable.

Third Semantic Attempt: Inferential Roles, Syntax, and Carnapian Tolerance

Let's look next at a rather different way of giving a semantic theory: via an inferential role semantics that associates each concept with governing inference principles. Given the Carnapian origins of the kind of conceptual maximalism that has spawned this investigation into the Boundary Question, perhaps we should expect this kind of syntacticized approach to language to bear fruit. Carnap, in his Principle of Tolerance, does seem to be treating it as a solution to the Boundary Question:

In logic, there are no morals. Everyone is at liberty to build his own logic, i.e. his own form of language, as he wishes. All that is required of him is that, if he wishes to discuss it, he must state his methods clearly, and give syntactical rules instead of philosophical arguments.

(Carnap 1934)

If we think of a language as a set of syntactic rules, then it looks like we have an easy answer to the question of what the possible range of languages is: we simply consider all possible syntactic rules. (Fussy point: we will need to take syntactic expressions to be individuated by their governing rules, so that we aren't at risk of overloading any one syntactic item. DWARF PLANET obeys a rule allowing transition from PLUTO IS A DWARF PLANET to PLUTO ORBITS A POINT OUTSIDE ITS BOUNDARY. PRIMARY PLANET obeys a rule allowing transition from JUPITER IS A PRIMARY PLANET to JUPITER ORBITS A POINT INSIDE ITS BOUNDARY. Those syntactic rule facts suffice to tell us that DWARF PLANET and PRIMARY PLANET are different syntactic items (and would be, even if both had the morphology "planet").)

There is a less tolerant and a more tolerant version of this Carnapian tolerance. On the less tolerant version, we require that a language be a collection of syntactic transformation principles that amount to rules of inference. On the less tolerant version, a "language" that consisted only of a single rule allowing "Jupiter" to be replaced by "Neptune" wouldn't count as a real language, because that rule wasn't allowing us to *genuinely infer* anything. Answering the Boundary Question in the less tolerant framework then requires a prior account of what makes a

syntactic transformation rule a genuine rule of inference, and this looks to be just as buck-passing as options we've already considered. (There won't be any progress, e.g., in saying that a syntactic rule is a rule of inference if it is appropriately truth preserving.)

On the more tolerant version, a language can be just any collection of syntactic transformation rules. The worry now is a new one: that the resulting framework is *too* generous, and that no sensible theoretical project can be maximal with respect to it. One version of this worry, of course, is the familiar "tonk" worry. There is a perfectly well-formulated syntactic rule for "tonk", so if our theorizing framework needs to be maximal with respect to syntactic rules and contain every expression for which a rule can be given, it must contain "tonk". But if we theorize using "tonk", then our theory will contain *everything*, and a norm of theorizing that requires us to end up with a theory containing everything can't be giving us a good picture of the theoretical enterprise. (In fact, things are even worse than that. We could also have a term "tunk" which is governed by the rule that it is *impossible* to move from "A tunk B" to "B tunk A". If we try to include both "tonk" and "tunk", then it's not just that our endorsed theory is inconsistent, but that we have inconsistent verdicts on what we should endorse. But it's unclear why rules of forbidding should be less acceptable than rules of permitting.)

There are well-known avenues for dealing with the "tonk" problem, of course—variants of constraining tolerance to some variant on conservative syntactic rules. But these avenues aren't terribly well-suited for maximalist enterprises. It's not hard to craft a pair of expressions such that adding either one to a core language is a conservative extension, but adding both is a non-conservative extension. (From A we infer "A tank B", and from A we infer "A tink B". From "A tank B and A tink B" we infer B, but there is no independent elimination rule for either "tank" or "tink".) Then we're left with no suitable maximal language.

One point that comes out from consideration of these heavily syntacticized options is that the specifically *alethic* formulation of the non-pragmatic norm of theorizing may need to go. This point lurked already in the earlier observation that our language might contain resources whose contents are given in terms of truth. When the applicability of truth gives out before the boundaries of the language, we need some more general picture of what our theorizing enterprise is beholden to—some notion, perhaps, of what is to be said in the language.

7. Answering the Boundary Question with Metasemantics

There are, of course, many more semantic frameworks from which we could attempt to extract an answer to the Boundary Question. But I think the pattern of difficulties we saw in the two cases above is likely to continue. In both cases boundary concerns about language/concept/proposition just get transferred over to the central semantic coin of the specific semantic framework being considered (possible world, truth), and no progress is made. Simply putting more coin on the table seems unlikely to change that difficulty.

Maybe, then, we need to look past our semantic theories to the metasemantics. Perhaps understanding *why* our semantic frameworks are using the semantic coins that they are will give us a better understanding of what the boundaries of those coins need to be. Note that we can distinguish two types of metasemantics. There is *grounding metasemantics*, which answers the question 'In virtue of what do expressions have the meanings that they do?'. Grounding metasemantics isn't obviously helpful for answering the Boundary Question, since the Boundary Question isn't concerned with the expressing relation between syntax and semantics, but just with the *range* of that relation. (It's not impossible that understanding the grounding of the expressing relation will yield insight into its range, of course.) And there is *guiding metasemantics*, which answers the question 'What theoretical role is being fulfilled by semantic values being what they are?'. It's the guiding metasemantics that's potentially helpful in answering the Boundary Question.

We'll consider three kinds of guiding metasemantics:

- Metaphysical guiding metasemantics, holding that the theoretical role of propositions is to stand in a representation relation to privileged 'sentence-shaped chunks of the world' (Rorty 1986).
- Cognitive guiding metasemantics, holding that the theoretical role of propositions is to be the possible contents of thoughts.
- Practical guiding metasemantics, holding that the theoretical role of propositions is to characterize a possible move in a communicative exchange.

First Metasemantic Attempt: Facts and Other Metaphysics Heavy-Weights

Suppose our metaphysics hands us a domain of facts, and contents are then determined by the facts (e.g., contents are just possible facts, so that the truth conditions of contents are those facts. But the details won't matter). Or instead our metaphysics might hand us a domain of states of affairs, or of fundamental entities and features—anything to which the theory of content is then taken to be representationally responsible.

The basic dialectic should be predictable at this point. Looking to the metaphysics for an answer to the Boundary Question threatens to leave us subject to a dilemma between **Undue Expressive Limitations** and **Passing the Buck**. Roughly, if the metaphysics is not providing us with what are stipulatively well-suited to propositional expression, then there are expressive limitation worries as we try to fit the theory of content onto what the metaphysics does give us. And if the metaphysics does give us sentence-shaped chunks of the world, there is a concern that we've just passed the buck to the metaphysics, and don't know any better how to answer a metaphysical analog of the Boundary Question.

Suppose what the metaphysics gives us is Lewisian possible worlds. Then we can avoid buck-passing worries, because we're in a good(ish) position to say what kind of thing a Lewisian possible world is (a maximally connected spatiotemporal region and its contents), and we can just let the world answer for how many and how diverse such regions there are. But if the metaphysics gives us Lewisian possible worlds, the semantics we can build off of it is a possible worlds semantics, and we're back with the **Undue Expressive Limitations** worries we encountered earlier for possible worlds semantics.

Suppose instead that the metaphysics gives us facts. Facts look much better (although not unquestionably perfect) for grounding a suitably expressive theory of contents. But now the question *what possible contents are there* just gets shifted to the question *what possible facts are there?* Is there a fact that Jones is a freethinker? Is there a fact that Sherlock Holmes is a detective? More pressingly, does the combination of, say, Pluto and Neptune constitute a fact? Or do we need to say what the **combination** relation is to answer this question? If so, the need to select among candidate combination relations threatens to be as hard as the question of selecting among candidate boundaries of concepts.

Second Metasemantic Attempt: Concepts as Thinkables

If *what a language is* is a tool communication of our mental states, then the starting point for answering the Boundary Question is thinking about our mental lives. What a proposition is, fundamentally, is the kind of thing that can be the content of one of our beliefs (etc.). To answer the Boundary Question, then, we need to start by figuring out what kinds of beliefs it's possible to have.

But this, of course, threatens to be buck-passing again. As usual, the worry can be posed using any of our stock supply of "defective concepts". But let's cut more directly to the core. Suppose we're presented with a creature and told that the content of one of its beliefs is SOUTH AMERICA or BEING A FAN OF BORGES. Can we give any convincing explanation of why those aren't possible belief contents? We might say that the regulative norm of belief is truth, and that SOUTH AMERICA, lacking truth conditions, can't be subject to such a regulative norm and hence can't be a belief. But this looks like we're just passing the buck one more step back to the theory of truth. If we can't answer the Boundary Question for truth conditions, then we can't (on this approach) answer the Boundary Question for beliefs, and thus can't answer the Boundary Question for contents. Or we might say that beliefs are teleofunctionally given states whose role is to track the facts—but then we've passed the buck back to the metaphysics. Or we might be functionalists about beliefs—but then we've passed the buck onward to the theory of action, and it's hard to see why we get an improved grip on the Boundary Question there.

Third Metasemantic Attempt: Moves in a Language Game

One last attempt. Perhaps the lesson of all of this is that the pragmatist was on the right track all the time—we need to give up on the "purely theoretical" project that inevitably begins by carving out a notion of content on theoretical metasemantic grounds, and then uses that notion of content to set out a domain of theorizing. Instead, we need to get our grip on the very *tools* of theorizing in pragmatic terms. We need to think about what a language *is* by thinking first about what a language *does*. On this view, a proposition is a device for producing certain kinds of effects.

The crucial question will then be: what kind of effects? If we try to set out the kind too narrowly, we're back in the soup above. (Consider: if we say that a proposition is a device for producing the effect of getting the audience to believe something-or-other, in Gricean metasemantic style, we're back in cognitivist metasemantic territory, and unless we can say what the range of believables is, we've made no progress.) If we set out the kind too broadly, we don't effectively carve out a propositional, or

even a linguistic, kind. (An utterance of "Aristotle" is going to do *something*, but that isn't enough to give that utterance propositional content.)

I think the best hope here is to appeal to the idea of "making a move in a language game". Consider the way that Dummett argues for a Fregean Context Principle in which sentential meanings have conceptual priority over subsentential meanings:

If I take some coloured counters, and say, 'Let this one stand for the Government, this one for the Opposition, this one for the Church, this one for the Universities, this one for the Army, this one for the Trade Unions, . . . ', and so on, I shall be understood on the presumption that I am about to make some arrangement of the counters by means of which I intend to represent some relations between these institutions, and assert that they obtain. If I do not go on to make any such arrangement, but simply start talking about something else, my earlier declarations lose their original intelligibility: I cannot, when questioned why I said all that, reply, 'Oh, I just wanted those counters to stand for those things, that's all'; for their standing for those things only amounts to anything if they are to be used to effect some symbolic representation by means of which a thought is expressed. Otherwise, my stipulation of their reference is like my saying, in the course of explaining a card game, 'Ace is high', and it later turns out that the ranking of the cards plays no role in the game; or it is like my saying 'Suppose there is life on Mars', and then failing to draw any consequences from this hypothesis, and, when challenged, saying, 'Oh, I simply wanted you to suppose that'.   (Dummett 1973)

But in the end, I don't find the suggestion helpful—I don't have any better grip on what it is to "make a move in a language game" than I do on the target notion of a proposition. I *do*, for example, find it intelligible to say "I just wanted those counters to stand for something". That in itself means that my notion of language game doesn't helpfully distinguish a 'claiming' move from a 'referring' move. That's enough to create worries for this pragmatic route to an answer to the Boundary Question, but it's also indicative of a deeper problem. In Dummett's relatively prosaic game, we can perhaps usefully distinguish claiming from referring moves, even if we regard both as legitimate moves. But as the games get more exotic, it's not at all clear that we have any helpful grip on what distinguishes an interesting *kind* of move.

8. Primitivism and the Star Gambit

Perhaps, of course, what all of this shows is that there's no *illuminating* answer to be given to the Boundary Question. Maybe all we can say is that there is a property of *being a language* or *being a proposition*, but we then have to take that property as primitive, and can't learn more about it by linking it to other properties (truth, fact, belief, action) in the way we've been attempting above. Or maybe it does link to some of these other properties, but in a way that produces a small and unilluminating circle.

It's hard to know what to say in response to "it's a primitive" moves. But I do think there is a specific and a general issue here each of which needs to be grappled with. The specific issue is that, as we've seen, there are particular questions about where the boundary lies that we'd like answers to. Does global language theorizing entail a theoretical commitment to catalog freethinker claims and tonk claims? The primitivist non-response to the Boundary Question leaves us at sea in answering those

particular questions. Of course, no one promised we'd get all of our questions answered. Nevertheless, these particular questions do seem like ones where we might have plausibly hoped for some helpful guidance.

The general issue is this: suppose we take the notion of a proposition to be a primitive. Nevertheless, surely in the end there are variant notions that are in some sense *proximate* to the notion of a language/concept/proposition, in the familiar "quantifier variance" manner. So in addition to languages/concepts/propositions, there are languages*/concepts*/propositions*. Then two closely connected questions:

1. In virtue of what are we talking about languages, rather than about languages*?
2. Why should we take the aim of theorizing to be to collect claims from all languages, rather than from all languages*?

In response to the first question, we can hope that there is some kind of metasemantic story that gets us pinned down to LANGUAGE, rather than to LANGUAGE*. As with familiar worries about the metasemantic commitments of epistemicism, we might worry about whether there is going to be a metasemantics that's precise enough to target just one out of a dense cloud of related notions. But even if we think that there is a precise metasemantics (maybe it's naturalness to the rescue), it's unclear whether it really helps. After all, if it's naturalness that targets LANGUAGE, there is presumably also NATURALNESS* targeting LANGUAGE*. Naturalness* might not be the source of our metasemantics, but it might be the source of our metasemantics*. Our metasemantics* doesn't determine what language we speak, but it might determine what language* we speak*.

Similarly in response to the second question. Suppose we get some convincing answer to why we ought to theorize in languages, rather than in languages*. Does this help? Maybe all it shows is that we ought to theorize* in languages*. But then we can suggest that, although that's true, we don't *care* about theorizing*—theorizing* isn't one of our projects. The counter-response, of course, is that although theorizing* isn't one of our projects, and isn't normatively important, it is one of our projects*, and is normatively* important*. The crucial question, I think, is whether the star gambit is revealing a *problem*, or is simply allowing us to say at each stage that we care about what we care about. I don't know the answer to that question.

References

Carnap, Rudolf. 1934. *The Logical Syntax of Language.* London: K. Paul, Trench, Trubner & Co.

Dummett, Michael. 1973. *Frege: Philosophy of Language.* London: Duckworth.

Hirsch, Eli. 2002. Quantifier Variance and Realism. *Philosophical Issues* 12 (1): 51–73.

Lewis, David. 1983. New Work for a Theory of Universals. *Australasian Journal of Philosophy* 61 (4):343–77.

Putnam, Hilary. 1978. There Is at Least One A Priori Truth. *Erkenntnis* 13 (1):153–70.

Quine, W.V.O. 1948. On What There Is. *Review of Metaphysics* 2 (5):21–38.

Rorty, Richard. 1986. Pragmatism, Davidson, and Truth. In E. Lepore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson* (pp. 333–5). Cambridge: Blackwell.

Sider, Ted. 2012. *Writing the Book of the World.* Oxford: Oxford University Press.