Higher-order metametaphysics*

Theodore Sider October 13, 2021

Abstract

Is higher-order quantification legitimate? I understand the question metaphysically, as asking whether reality is such as to be well-represented by such languages, so that their sentences do not lack truth value, or determinacy, or objectivity, or suffer any other sort of "discourse failure". I critically discuss three arguments for the legitimacy, thus understood, of higher-order languages: that the languages are needed for a proper statement of set theory, that they are needed for a proper account of semantics, and that they can be used to give a reductive theory of necessity.

The recent turn to higher-order languages—languages with quantification into predicate, sentence, and other nonnominal positions—promises elegant and more accurate modes of expression, new solutions to old problems, transformation of old problem spaces, and generation of new questions: a paradigm shift. The excitement in the peroration of Cian Dorr's agenda-setting paper "To be *F* is to be *G*" is typical of the spirit of this movement, and undeniably infectious:[1]

> And the exploration has barely begun: there is a whole continent of views waiting to be mapped out, and at this point we can only guess which of them will look most believable in the long run. Onwards!

As a community, the best way to handle such new, all-encompassing, and programmatic ideas is to run with them. Many of us should embrace the new framework, explore it from the inside, and see where that leads. Setting aside the inevitable nay-sayers, that's what we did with Frege's logic and linguistic analysis in the early twentieth century, and with possible worlds and modal logic in the 1970s (to take just two examples), in each case with great success. In-depth exploration is needed to tell whether ideas are on the right track; we know them by their fruit.

But nay-saying has its place too. Strawson and the other ordinary-language philosophers provided a corrective to Russell and his heirs, important parts of

*A more accurate title would be "Meta-(higher-order metaphysics)", but.... Thanks to Jeff Russell for helpful discussion.

[1] Dorr (2016, p. 84). Other literature includes...

which were eventually assimilated into the mainstream. Quine (at the very least) forced modal enthusiasts to clearly articulate and embrace their metaphysical commitments.[2] And sometimes nay-sayers are right. The community needs them too.

Despite its undeniable appeal and promise, there are important metaphysical questions about the foundations of the higher-order approach. Are higher-order languages in metaphysically good standing? That is, do such languages succeed in latching onto reality; is reality such as to be well-represented by them? If so, then it would indeed make sense to stay up nights wondering whether, for instance, $\forall p\, p = p \,\&\, p$. Such questions would concern reality's higher-order structure. If not, the questions might not have answers at all (if higher-order sentences fail to be truth-apt), or might fail to have determinate or objective answers, or might have "unwanted" answers (if, say, higher-order sentences have first-order, set-theoretic truth conditions), or might suffer some other sort of "discourse failure".[3]

1. Higher-order languages

In the late nineteenth century, the concept of set—the concept of a collection conceived as an individual thing—became central to the foundations of mathematics. But around the turn of the century, apparent contradictions in this idea were discovered, the simplest and most famous of which is Russell's. Define r as the set of all and only those sets that are not members of themselves. Thus a set is to be a member of r if and only if it is not a member of itself:

$$\forall x(x \in r \leftrightarrow \sim x \in x) \tag{R}$$

[2] Eli Hirsch (2011) and Amie Thomasson (2007, 2015) have played a similar role in another context, forcing the dominant "Quinean" tradition in ontology to articulate and defend their foundational assumptions.

[3] One might stay up nights even given discourse failure. Answers are not the only value in questions. In any inquiry into metaphysics (or anything in philosophy, for that matter), there is also the value of cartography of logical space, including the creative/expansive invention of new positions, and the dialectical/contractive search for reasons against positions. Because higher-order metaphysics is so formal, cartography in this domain has a certain kind of known value, the same value that attaches to mathematical investigations independent of their truth. But cartography has value even in less formal metaphysics, where we rely less on the kinds of clear-cut considerations logic studies, and more on the woollier, less well-understood (but no less essential) considerations we call philosophical. (The latter are still needed in higher-order metaphysics, for instance in informing the ubiquitous judgments about which hypotheses are worth exploring and which are not.)

As Russell observed, if we substitute 'r' for the universally quantified variable 'x', we obtain:

$$r \in r \leftrightarrow \sim r \in r$$

which is a contradiction.

Two parts of this reasoning bear emphasis. First, we are treating terms referring to sets as being grammatically like terms referring to their members. Thus in addition to formulas like $x \in r$, formulas like $x \in x$ and $r \in r$ (and thus their negations $\sim x \in x$ and $\sim r \in r$) are also grammatical. In modern terms, we are speaking of sets using a first-order language; we refer both to sets and their members using singular terms; and we ascribe membership using a two-place predicate $\in$.

Second, the existence of the set r is simply assumed. Without that assumption there is no paradox, just as the "paradox" of the barber who shaves all and only those who don't shave themselves is easily dissolved by denying the existence of the barber. As we now put it, Russell's paradox refutes the assumption of Naive Comprehension, according to which every formula determines a set. Naive Comprehension is this schema:

$$\exists y \forall x (x \in y \leftrightarrow \phi) \qquad \text{(Naive Comprehension)}$$

where ϕ may be replaced with any formula with no free occurrences of variables other than x.[4] Replacing ϕ with the formula $\sim x \in x$ yields the instance $\exists y \forall x \big(x \in y \leftrightarrow \sim x \in x\big)$; existentially instantiating to an arbitrary name r yields the contradictory sentence (R).

The dominant approach to the paradox has been to reject the second assumption (and thus Naïve comprehension). The set r doesn't exist. We cannot simply assume that every formula corresponds to a set. Rather, we must carefully develop a theory of when sets exist and when they do not, a theory which implies the existence of all the sets we need in mathematics but does not imply contradictions like (R). Zermelo Frankel set theory is an elegant theory of this sort, and is the dominant theory of sets today.

But there is another possible approach: reject the first assumption, according to which a term for a set and a term for one of its members have the same grammar. Russell and Whitehead adopted such an approach, known as the ramified theory of types, in *Principia Mathematica*. That theory was ungainly

[4]This is a special case. The more general version of the schema allows "parameters": instances may be prefixed with any number of universal quantifiers binding variables which may occur freely in ϕ.

and soon became obsolete, but an improved version due to Church (1940) and others, known as the simple theory of types, continued to be studied by logicians (and its descendants by computer scientists); and it is this and related theories that have become so popular in recent metaphysics, philosophy of language, and philosophical logic.

Actually this type-theoretic approach has been used to develop a consistent theory of properties, relations, propositions, and the like, rather than sets. These entities are capable of playing a similar role to sets in the foundations of mathematics.[5] And inconsistency threatens these entities just as it threatens sets: uncritically assuming the existence of a property for each predicate yields the property of not instantiating itself, which would then instantiate itself if and only if it does not instantiate itself. The type-theoretic resolution of this paradox is roughly that an expression for a property is not grammatically like an expression for one of its instances, so that a statement saying that a property instantiates itself will be ill-formed, and the paradox does not get off the ground.

In more detail: the simplest resolution of the paradox takes talk of properties to be formalized using the language of second-order logic, in which variables whose grammar is that of predicates are allowed, in addition to variables like those of first-order logic, whose grammar is that of names. Thus in second-order predicate logic (but not first-order logic) formulas such as these are well-formed:

$$\exists F F(a)$$

$$\forall R\big(R(a,b) \to R(b,a)\big)$$

$$\forall x \exists F\, F(x)$$

"Quantification over properties" is thus understood as quantification into predicate position: $\exists F$ and $\forall F$. And the "attribution of properties to things" is achieved, not by interposing a predicate of instantiation between a singular term naming the thing and a singular term naming the property, as in 'x instantiates y' (compare '$x \in y$') but rather by attaching a predicate to a singular term: '$F(x)$'. The attempt to formulate a claim that a property instantiates itself then becomes '$F(F)$', which is as ungrammatical in second-order logic as it is in first.

The language of second-order logic is a special case of the higher-order languages that are now popular, which in general go beyond second-order languages in two ways: they allow constants and variables of arbitrary grammatical categories, and they allow lambda abstraction.

[5] Note

The notion of "arbitrary grammatical categories" is made precise by the device of *types*. Types are conventional entities used to represent, or code up, grammatical categories; thus we speak of expressions in formal languages as *having* or being *of* types. The purpose of representing grammatical categories as entities is to allow us to quantify over them in the metalanguage, in order to make generalizations about expressions of different grammatical types: "for any type, τ, if an expression has type τ, then ...".

Here is one typical development of the idea. We begin with two types, t and e. What *are* these types? It doesn't matter; the association between these entities and grammatical categories is purely conventional. We might as well take them to be the *letters* 't' and 'e'. t represents the grammatical category of expressions that are capable of having a *t*ruth value (i.e., formulas) and e represents the grammatical category of expressions that stand for *e*ntities (i.e., singular terms). e and t are called "primitive" types because they aren't constructed from any other types. All the other types are constructed from simpler types, according to this rule:

$$\text{if } \tau_1, \ldots, \tau_n \text{ are types, then } (\tau_1, \ldots, \tau_n, t) \text{ is also a type} \tag{T}$$

(Again, keep in mind the conventionality of types. $(\tau_1, \ldots, \tau_n, t)$ can be regarded as nothing more than the left paren, followed by τ_1, followed by a comma, followed by τ_2, then a comma, ..., followed by τ_n, followed by the letter t, followed by the right paren.) $(\tau_1, \ldots, \tau_n, t)$ represents the grammatical category of an expression that combines with n expressions, of types $\tau_1, \ldots, \tau_n$, respectively, to form an expression of type t, i.e., a formula. For example, an expression of type (e, t) combines with an expression of type e (a singular term) to form an expression of type t (a formula). That is, it's a one-place predicate. And an expression of type (t, t, t) combines with two expressions of type t (i.e., with two formulas) to make an expression of type t (a formula). That is, it's a two-place sentence operator, such as $\wedge$ or $\vee$.

The rule (T) can be applied iteratively, since $\tau_1, \ldots, \tau_n$ may be any types, including complex types. Since (e, t) is a type, so is $((e, t), t)$; but then $(((e, t), t), t)$ is also a type; and so on. There are infinitely many types.

In a higher-order language based on this simple type theory, constants and variables of each of the infinitely many types are allowed. Thus in addition to quantifying into singular-term-position (as in first-order logic), or predicate position (as in second-order logic), one can quantify into sentence position (variable of type t), as in:

$$\forall p \exists q (q \leftrightarrow \sim p)$$

("for every proposition, there exists a proposition that is true iff the first is not true"), or into one-place sentence-operator position:

$$\exists O \forall p (O(p) \leftrightarrow \sim p)$$

("There exists an operation that, when applied to any proposition yields a proposition that is true iff the first proposition is not true"), or any other position represented by a type.[6]

In addition to quantification into positions of all types, the currently popular higher-order languages include a second innovation (also due to Church): lambda abstraction. The purpose of lambda abstraction is to allow for complex expressions of arbitrary (complex) type. For example, in addition to a simple predicate S for 'sits' and a simple predicate E for 'eats', we might want a complex predicate meaning 'sits-and-eats'. (Why? For one thing, to be an allowable substitution for predicate variables in quantified sentences. 'For all properties, if John has the property then Ted has the property' should imply something like 'If John sits-and-eats then Ted sits-and-eats'.) We represent 'sits-and-eats' as $\lambda x.(Sx \,\&\, Ex)$, read as "is an x such that x sits and x eats". In general, where $v_1, \ldots, v_n$ are any variables, of types $\tau_1, \ldots, \tau_n$, respectively, and ϕ is any formula, then $\lambda v_1, \ldots, v_n.\phi$ is an expression of type $(\tau_1, \ldots, \tau_n, t)$, meaning "are $v_1, \ldots, v_n$ such that ϕ".[7]

2. "Innocent" higher-order quantification

Quine famously said that second-order logic is set theory in sheep's clothing, meaning that a second-order sentence like $\exists F F(x)$ really just means that x is a member of some set. (Or that x instantiates some property; but Quine took talk of properties to be strictly less clear than talk of sets.) That is, a more perspicuous way of saying what you were trying to say would use a first-order language: $\exists y\, x \in y$ (or: $\exists y\, I(x,y)$—I for "instantiates"). More perspicuous because it makes clear our ontological commitments: by saying it in the first-

[6] Here I am using p and q as variables of type t, and O as a variable of type (t,t). Often the types of expressions are represented explicitly by superscripting: p^t, q^t, $O^{(t,t)}$.

[7] A nice perk of lambda abstraction is that it can take over the job of variable binding from quantifiers; quantifiers no longer need to be treated as syncategorematic. For instance, we can replace $\forall x F(x)$ and $\exists x(F(x) \,\&\, G(x))$ with $\forall F$ and $\exists \lambda x.(F(x) \,\&\, G(x))$, in which $\forall$ and $\exists$ are treated as expressions of type $((e,t),t)$, attaching to predicates (like F and $\lambda x.(F(x) \,\&\, G(x))$) to form sentences.

order way, we no longer hide our commitment to an ontology of sets (or properties).

The central presupposition of higher-order metaphysics is that Quine is mistaken about this. To be sure, one *could* decide to use a higher-order language to express the same claims that might otherwise be expressed in a first order language quantifying over sets or properties. But it is also possible to take the higher-order quantifiers as sui generis, to take higher-order sentences to express claims that cannot be expressed in first order languages. Such claims do not involve quantification over sets or properties "as entities", by which we mean simply that there need not exist—first-order quantifier—any sets or properties in order for existential higher-order claims to be true. $\exists F\, F(x)$ does not mean that there exists (first-order quantifier) some entity that x instantiates or is a member of. It can be true even if there are no such things (first-order quantifier) as sets or properties. So what does it mean, then? It means, well, that $\exists F\, F(x)$. Similarly for other types. $\exists p\, p$ does not mean that there exists (first-order quantifier) some proposition that is true; it can be true even if there are no such things as propositions. It means, well, that $\exists p\, p$.

If this view is correct, then many of the intuitive glosses of higher-order claims that I have been giving (and will continue to give) are misleading. Strictly speaking we should not gloss $\exists F\, F(x)$ as "x has some property", or $\forall p\, p = p \,\&\, p$ as "every proposition is identical to its self-conjunction", since each gloss suggests first-order quantification, over properties in the first case and propositions in the second. Indeed, it's not clear that higher-order claims can be perspicuously stated in natural language at all.

This anti-Quinean view is sometimes put by saying that higher-order quantifiers are "ontologically innocent", that they are not "ontologically committing"; but this can seem to mean more than it does. What it does mean is that higher-order quantification does not commit us to entities *in a first-order sense*. $\exists F\, F(x)$ can be true without there being some entity (first-order quantifier) corresponding to the predicate variable F. Nevertheless, there is a perfectly good sense in which it is "ontologically committing". $\exists F\, F(x)$ is, after all, an existential sentence, and says that *there is* an F of a certain sort; it is false if there is no F that x has, in the second-order sense of 'there is no'.

George Boolos (1984) famously defended an anti-Quinean view in this vicinity, according to which *plural* quantifiers, such as 'some' in 'some pallbearers lifted the casket', are sui generis, and not first-order quantifiers over sets or the like. Although the following discussion is meant to encompass Boolos's view, the current higher-orderists who are my main target depart from it in two ways. First, while Boolos gives us a way of interpreting the language of <u>monadic</u>, <u>second-order</u> logic (in which the only non-first-order quantification is into the position of one-place first-level predicates, i.e., type (e,t)), the authors I have in mind embrace quantified variables of arbitrary type, and thus of arbitrary 'adicy and level. Second, plural variables are "extensional": given the plural interpretation, the following sentence is true:

$$\forall F \forall G\big(\forall x(Fx \leftrightarrow Gx) \rightarrow F = G\big)$$

If every creature with a heart is a creature with a kidney, and vice versa, then the creatures with hearts (plural variable) *are* the creatures with kidneys. The higher-orderists, on the other hand, do not accept the sentence. Thus their "$\forall F$" is more akin to "all properties" than "all pluralities" (setting aside the misleading suggestion of first-order quantification over properties).

In the claim just displayed, an identity predicate was flanked by second-order variables. Such higher-order identity predicates play a central role in many of the higher-orderists' inquiries. Suppose that, for any type, τ, we introduce an identity predicate for that type, $=^\tau$. We may then raise the question of "fineness of grain" for type τ: under what conditions are "entities of type τ" the same or different? For instance, are propositions (forgive the first-order sound) individuated by truth value (the coarsest imaginable grain)? The thesis that they are may be stated as follows:[8]

$$\forall p \forall q\big((p \leftrightarrow q) \rightarrow p =^t q\big) \qquad \text{(extensional propositional grain)}$$

Are they instead individuated by necessary equivalence? Helping ourselves to an operator $\Box$ for necessity, that claim would be:

$$\forall p \forall q\big(\Box(p \leftrightarrow q) \rightarrow p =^t q\big) \qquad \text{(intensional propositional grain)}$$

Or perhaps they are even finer-grained? Since $\Box(p \leftrightarrow (p \,\&\, p))$, intensional propositional grain implies:

$$\forall p\; p =^t p \,\&\, p$$

which is incompatible with a competing idea about propositional grain: that propositions are "structured".

[8]'$\leftrightarrow$' is the material biconditional; $\phi \leftrightarrow \psi$ is true if and only if ϕ and ψ have the same truth value.

Similar questions of grain may be raised for any type, with the help of lambda abstraction. Are properties identical to their self-conjunctions: $\forall F\, F =^{(e,t)} \lambda x.(Fx \,\&\, Fx)$? Is negation the same as triple negation: $\sim =^{(t,t)} \lambda p.(\sim\sim\sim p)$? And so on.

Claims of all these sorts are of course ungrammatical in first-order logic. Adopting the higher-order language opens up Dorr's continent of possible views about grain.

So: is higher-order logic "innocent"? Better: should we accept irreducibly higher-order quantification?

It isn't fully clear what Quine himself meant by calling second-order logic set theory in sheep's clothing, or what his reasons were.[9] Sometimes he simply assumes, begging the question, that all quantified variables range over entities (Quine, 1970, pp. 66–7). Sometimes he is insisting that second-order logic is no more part of logic proper than first-order set theory; but I don't read the current higher-orderists as denying this (or even regarding logicality as an important classification).

In my view, the prima facie case against accepting irreducibly higher order quantification is simply parsimony. Posits which make the world more complex are, other things being equal, to be avoided. And the posit of higher-order quantification makes the world much, much more complex. Dorr's continent, exciting as it admittedly is, is exactly the problem. The posit commits one to an ocean of new facts, and the size of the continent is a testament to the size of the ocean.

When I embraced the logical apparatus of first-order logic, for instance using an embeddable negation sign, I didn't sign up for questions such as whether $\sim =^{(t,t)} \lambda p.(\sim\sim\sim p)$. The higher order framework is a massive, massive jump in expressive power, leading to the recognition of a host of new facts. This jump in expressive power does indeed lead to exciting opportunities for new research, but the downside is a great increase in the complexity of the world. This is a very basic and common sort of recoil from a proposed metaphysical commitment.[10] The added complexity might be worth it, just as the added worldly complexity required by the posit of properties in physics such as charge is presumably worth it, but is nevertheless a cost, not to be paid lightly.

[9] See Boolos (1975) and Turner (2015).

[10] Another sort is more common but less defensible I think: recoil from apparently unknowable facts. The epistemic recoil has a quite different source, neo-verificationist rather than Occamist. See Sider (2020, section 3.15) for a discussion of related issues.

In the sequel we will examine some arguments that the cost is worth paying.[11] But it may be objected right at the start that there is no cost at all, precisely because higher-order quantification is innocent.

I have already indicated in a preliminary way my objection to this thought: although higher-order quantifications are not ontologically committal in a first-order sense, they are ontologically committal in a higher-order sense. But more can be said.

First, many sorts of "metaphysical commitment" involve no first-order ontological commitments. For instance, the adoption of a modal outlook—using modal operators in our theorizing—presumably has no distinctive ontological commitments: the modal operators do not correspond to new entities, but rather "new modes of truth", so to speak. For "modalists", reality has a modal aspect, an aspect unrecognized by anti-modalists like Quine. Modalists accept an ocean of facts, resulting in a continent of new questions, such as whether reality might have been exactly as it actually is physically but not mentally, whether I could have been born from different parents, and so on. The world is a far more complex place for them than it is for Quine, despite the fact that they don't (or needn't) recognize any new entities.[12] Similarly, the adoption of predicates of 'charge' and 'mass' in physics, which everyone acknowledges as involving an increase in worldly complexity, don't (or needn't) involve postulation of new entities.

In each case, one might argue that new entities should indeed be postulated: universals of charge and mass, properties of necessity and possibility possessed

[11]Could the continent itself justify the cost? The ideological and ontological posits of medieval angelology are hardly justified by its continent of questions. But investigation of the higher-order continent is more formally disciplined than inquiry into angels dancing on pin-heads. In the investigation of Dorr's continent that is now underway, there is a sense that we are getting some formal traction: the investigation can be carried out with mathematical rigor, certain initially natural-seeming ideas can be demonstrated to be inconsistent, formally natural groupings of viewpoints have begun to emerge, and so on. It's an interesting question whether this is any sort of evidence that we are making contact with reality. It's important not to confuse the admitted value of mathematically disciplined cartography of logical space with evidence of contact with reality. Still, some may claim that in pure mathematics itself, formal traction is evidence of contact with reality. The most extreme version of this view is one commonly associated with Hilbert (e.g., 1899), namely that consistency implies truth in mathematics; an ideological variant would say that consistent ideology is ipso facto in good standing.

[12]My quantification over aspects, modes, and facts in this paragraph is inessential, present only because of natural language's preference for nouns (illustrated at least twice by this very sentence).

by propositions. Whether there are such entities is a matter of controversy: "platonists" say there are, "nominalists" (like me) say there aren't. But it shouldn't be controversial that if nominalism is true, the adoption of modal operators or predicates of charge and mass still amounts to the recognition of added worldly complexity, of a sort that ought, other things being equal, to be minimized.

I myself think of the metaphysical commitment in a certain way: as a commitment to the expressions in question "carving at the joints" (Lewis, 1983*a*; Sider, 2011).[13] But the point here is not tied to this metaphysical baggage. Even those who are skeptical of it should, unless they reject realist philosophy of science in general, agree that the adoption of predicates for charge and mass is "costly" in the Occamist sense. And then, unless they claim some special exemption for metaphysics, they should agree that the adoption of modal operators is costly in the same sense. And then, unless they claim some special exemption for logic, they should agree that the adoption of higher-order languages is also costly.[14]

Objection: some of the central questions of higher-order metaphysics have to do with fineness of propositional grain, the question of when $p = q$, for arbitrary p and q. Some views about propositional grain would in a sense reduce the sizes of the ocean and continent: apparently different propositions about higher-order matters would in fact be identical. I have said that the higher-order viewpoint increases reality's complexity because, for example, there is then a fact of the matter as to whether (for instance) propositions are individuated by their structure, or by necessary equivalence, or by some other criterion. But it might be that propositions are in fact individuated coarsely enough so that, for instance, the proposition that propositions are individuated by necessary equivalence is *identical to* the proposition that propositions are individuated by their structure. And if so, it might be concluded, the adoption of higher-order languages doesn't add to worldly complexity after all.

This argument assumes that worldly complexity, of the sort relevant to theory choice, should be understood in terms of propositional fineness of grain, of the sort captured by higher-order propositional identities. I reject this assumption. Suppose the correct individuation of propositions is maximally coarse: they are individuated by truth value. That would trivialize all judgments of complexity whose locus is at the propositional level, since no matter what

[13] See also Fine (2001) for an approach in the same quadrant of logical space.

[14] Timothy Williamson forcefully opposes both exemptions, for instance in the preface to *Modal Logic as Metaphysics*. Other works in the anti-exceptionalist tradition include Almog (1989); McSweeney (2019); Paul (2012); Quine (1948).

there would be just two propositions, The True and The False.[15] Yet surely, even given the truth of a conception of propositional grain, complexity would still be a nontrivial theoretical indicator, in physics for example. I conclude that the relevant sense of complexity must be detached from the truth about grain.[16]

3. Argument from natural language

Some defend irreducibly higher-order quantification by arguing that natural language already contains it. Boolos (1984) famously argued that natural language contains plural quantification; and Agustín Rayo and Stephen Yablo (2001) (following Arthur Prior (1971) and Dorothy Grover (1992)) argue that natural language contains devices tantamount to both monadic and polyadic second-order quantification. Just as Boolos claims that the natural language sentence 'Some critics admire only one another' doesn't carry a commitment to sets of critics, so Rayo and Yablo argue that 'Somehow things relate such that everything is so related to something' (the putative natural-language analog of $\exists R \forall x \exists y R(x, y)$) doesn't carry a commitment to relations (as entities).

But it isn't clear why any of this matters. If natural language doesn't contain higher-order quantification, couldn't we just introduce it, provided reality can support such talk? Conversely, if natural language does contain higher-order quantification, but reality *can't* support it, wouldn't natural language higher-order claims then be subject to some sort of discourse failure, or ultimately be made true by purely "first-order" facts? The real issue is whether reality can support higher-order talk, not whether natural language already has it. Moral and modal skepticism of various sorts persist (including error theories, expressivist theories, and aggressively reductive theories) despite the presence of modal and moral natural language; why should matters be different with higher-order quantification?

The primordial issue, as I say, is the purely metaphysical one of whether reality can "support" higher-order languages. That is, does reality contain higher-order facts? Do irreducibly higher-order claims make adequate contact

[15] Similarly, a maximally coarse individuation of properties, namely by extension, would trivialize complexity judgments whose locus is at the level of properties.

[16] How *should* it be understood? This is a difficult question that I have no complete answer to. In light of the present considerations, it's natural to turn to a more syntactic approach, on which a theory's complexity has to do with the set of descriptions of the world that can be constructed using the vocabulary of the theory. But this skeletal beginning leaves many questions unanswered.

with reality? Now, it is difficult to express this issue in some canonical way that is both theoretically satisfying and neutral on certain questions of metaphysics. As in the previous section, one might understand the issue in terms of a metaphysics of carving at the joints.[17] But it's again important to see that such metaphysical baggage is not required. It is plain that there are analogous, gripping questions about both modal and moral language, and that those questions do not grip us solely because of an antecedent acceptance of inflationary metaphysics.[18] So let us continue with an atheoretical, baggage-free statement of the question: does higher-order talk have the underpinnings in reality needed to be free from either reduction or discourse failure of various sorts?

4. The ZF argument

Another of Boolos's arguments has nothing to do with natural language, and bears on the question of the standing of higher-order logic as I understand it.

The argument begins by assuming the correctness of the Zermelo-Frankel approach to set theory. (Thus the argument supports higher-order logic as a supplement to, not a replacement for, set theory. All it immediately supports, however, is plural quantification, or monadic second-order logic.) In its standard, first-order axiomatization, ZF set theory contains the following axiom schema:

$$\forall z \exists y \forall x \big(x \in y \leftrightarrow (x \in z \;\&\; \phi)\big) \qquad \text{(Separation)}$$

Despite its similarity to Naive comprehension, Separation does not imply Russell's contradiction. Whereas Naive comprehension says that any condition picks out a set, Separation says merely that any condition picks out a *subset* of any *given* set z. Substituting '$x \notin x$' for ϕ yields an instance saying that for any set z, there exists a set containing all and only the members of z that are not members of themselves. But this isn't contradictory: assuming that no set is a member of itself (which is guaranteed by another axiom of ZF), this subset will simply be all of z.

The idea behind ZF is to replace the "generation" of sets by Naive Comprehension with a two-step process of generation: there are "building" axioms (the null set, pairing, unions, and powerset axioms, and the axiom of infinity), which

[17] Or in terms of related ideology, as in Fine (2001).

[18] It is compatible with this methodological point that the best way of understanding the questions in fact makes use of inflationary metaphysics.

tell us that certain initial sets exist, and then there is the Separation schema[19], which lets us "carve away" arbitrary subsets of the initial sets.

But there is an apparent problem with this procedure: the Separation schema doesn't really imply the existence of "arbitrary subsets". Rather, it generates only those subsets that can be picked out by a formula, ϕ, in the language of set theory. Thus the infinitely many instances of the Separation schema, taken as a whole, fall short of expressing what one might have thought is the intuitive idea, namely that *any collection of members of a given set, z, forms a set*. Collections that are inexpressible by formulas are left out. (Since the language of set theory has an enumerable vocabulary and its formulas are finitely long, its set of formulas is enumerable, whereas the set of subsets of any infinite set is not enumerable.)

But what else could "any collection" mean? It can't mean "any set", since the comprehension principle would then become the trivial statement that any subset of a given set, z, forms a set. The higher-orderist has an answer: "any collection" can be understood in second-order terms: for any *property*, in a second-order sense of "any property", there exists a subset of any given set z consisting of z's members that have the property. Or, as Boolos himself argued, "any collection" can be understood in plural terms: for any things (universally quantified plural variable), there exists a subset of any given set z consisting of z's members that are among those things. Taking the former approach, we can replace the separation schema with a second-order separation axiom (not a schema, a single sentence):

$$\forall z \forall F \exists y \forall x \big(x \in y \leftrightarrow (x \in z \,\&\, F(x))\big) \qquad \text{(Second-order Separation)}$$

The argument in favor of higher-order quantification, then, is that with it we can better formulate the axioms of ZF set theory:

> It is, I think, clear that our decision to rest content with a set theory formulated in the first-order predicate calculus with identity... must be regarded as a compromise, as falling short of saying all that we might hope to say. Whatever our reasons for adopting Zermelo-Fraenkel set theory in its usual formulation may be, we accept this theory because we accept a stronger theory consisting of a *finite* number of principles, among them some for whose complete expression second-order formulas are required. We ought to be able to formulate a theory that reflects our beliefs. (Boolos, 1984, p. 441)

[19] Really, the Replacement schema.

Now, the alleged problem with first-order ZF cannot be that there are truths about sets that we can't state using its language. Assuming that *any* language we can speak has an enumerable vocabulary with finitely long sentences, there will always be too many truths about sets to state them all individually, no matter what language we speak.

The quotation suggests that the problem involves our *beliefs*: their statement requires expressive resources beyond first-order ZF. This is still weak: we might believe things that don't correspond to anything in the world, or are even incoherent. If the beliefs of some sect of medieval angelologists require distinctive ideology to state, this is no argument that the ideology corresponds to anything real.

An initially more promising argument is that second-order ZF is a *better theory*, and that this gives us a reason to accept any conceptual resources needed to state it.

Compare the posits we make in the physical sciences for the sake of theoretical gain. The best reason to believe that there are fundamental physical properties and relations is that only by positing them can we state the laws of dynamics. For instance, take charge. We observe regular patterns in the motions of things. When certain things get close to each other, they move apart—some quite sharply, others less so. When certain other things get close to each other, they move even closer together—some quite sharply, others less so. It is natural to posit that the particles have different properties, which we name charges; that there are laws relating force to charge; and that there are dynamical laws saying how things move as a function of the forces acting on them. Thus we posit properties like charge in order to be able to state strong laws of nature. If we didn't posit charge, we couldn't state laws of dynamics; we could only have list-like statements: "these particles moved in these ways, those moved in those ways, …".[20] Similarly, perhaps, we should treat higher-order quantification as a theoretical posit, like the posit of charge. Without it, we have only the list-like separation axioms of first-order ZF. With it, we can state the *law* of second-order separation.

The analogy isn't perfect, since the Separation axioms aren't really list-like: all instances of the schema share a certain syntactic form. The problem with stating laws of motion without properties like charge is that in the list of descriptions of particles' motions, no predicates beyond spatiotemporal ones occur, so there is no pattern; but in the list of instances of Separation, each

[20] See Sider (2020, sections 4.4 and 4.12) for a discussion of some related issues.

instance has a rich syntactic structure, in virtue of which it makes sense to say that each shares a single syntactic form. So the set of instances of the Separation schema is *not* a complex, miscellaneous set in the way that the list-like set of descriptions of particle motions is. It exhibits a simple pattern; it is unified by a single syntactic relation holding between all its members.

In fact, once we remember that our total theory includes laws of logic, not just laws of nature and laws of mathematics, it becomes clear that the sort of unification-by-pattern that we observed with the instances of the Separation schema is ubiquitous. The laws of propositional logic, for instance, are usually stated using schemas, such as:

$$\big(\phi \to (\psi \to \phi)\big)$$

That can be avoided by replacing the schema with a single axiom, involving particular sentences or sentential variables, such as 'Snow is white → (grass is green → snow is white)', and then adding a rule of substitution allowing us to infer a uniform substitution instance from any theorem (thus from the new axiom we may infer by substitution an arbitrary instance of the original schema, by replacing 'Snow is white' with ϕ and 'Grass is green' with ψ). But this just shifts the bulge in the carpet. For rules of inference themselves, even simpler ones like modus ponens, are not statements in the language of first-order logic, but rather are relations between formulas. The theorems of propositional logic form a simple, unified whole because of the patterns they collectively instantiate; and these patterns are not exhausted by a statement of the axioms that generate the whole, but in addition involve the fact that the set of theorems is closed under the rules. (This is Lewis Carroll's (1895) point.) Their collective simplicity involves the holding of the relations that are the rules.

So the mere fact that the axioms of first-order ZF must be stated schematically does not detract from its quality as a theory. Its theorems, as a whole, exhibit a pattern, in part captured by the fact that each instance of the Separation schema shares a single syntactic form; and this sort of pattern is akin to patterns that unify any theory closed under rules of inference.

The ZF argument should not be understood as targeting the simplicity of first-order ZF, but rather its *strength*. (This is clearly how Boolos understood it, though perhaps not with the spin I am giving it by analogizing to dynamical laws.) "Schematic laws"—that is, sets of statements unified by a syntactic relation—can be simple, as we've seen; but the instances of the Separation schema are collectively weaker than the second-order axiom of Separation. So the argument is this. A good theory should have laws that are both simple and strong; and although first-order ZF has laws that are simple (in some cases in the schematic sense that we have been discussing) and somewhat strong, it does not match second-order ZF's combination of simplicity and strength.

The argument can't be left there, as will emerge if we consider how Separation is applied. Let A and B be any two sets. A simple theorem of ZF is the statement that there exists such a set as the intersection of A and B—the set of things that are elements of both A and B. In first-order ZF, this is proven by beginning with this instance of the Separation schema:

$$\forall z \exists y \forall x \big(x \in y \leftrightarrow (x \in z \;\&\; x \in B) \big)$$

(which we obtain by changing the schematic letter ϕ to "$x \in B$".[21]) Then we let z be A, and infer:

$$\exists y \forall x \big(x \in y \leftrightarrow (x \in A \;\&\; x \in B) \big) \tag{1}$$

y is our desired intersection of A and B.

The proof in second-order ZF is a little different. We can begin by inferring the following from the second-order Separation principle (letting z be A):

$$\forall F \exists y \forall x \big(x \in y \leftrightarrow (x \in A \;\&\; F(x)) \big) \tag{2}$$

Next we need to instantiate the variable F to the property of "being a member of B". But how do we do that? The answer depends on what *logical* principles we take to govern the second-order quantifiers.

Typical axiomatizations of second-order logic include the following schema (where ϕ may be replaced with any formula with no free occurrences of variables other than x):[22]

$$\exists F \forall x (F(x) \leftrightarrow \phi) \tag{Comprehension}$$

[21]Really we must use a more general version of the subset schema that allows parameters—see note 4.

[22]And perhaps parameters; see note 4. An axiom schema of Comprehension is not needed in a language with λ abstraction, a rule of universal instantiation in which the universally quantified variable can be instantiated to λ abstracts, and the schema of "β conversion": $\lambda v.(A)\alpha \leftrightarrow A_v(\alpha)$, where $A_v(\alpha)$ is the result of changing vs to αs in A (in accordance with the usual restrictions). (To derive an arbitrary instance of Comprehension, begin with $\forall G \exists F \forall x (Fx \leftrightarrow Gx)$; infer $\exists F \forall x (Fx \leftrightarrow \lambda x.(A)x)$ by universal instantiation; then use β conversion to derive the instance.) But this wouldn't affect the argument, since the theory of properties would share the same expressive weakness as the theory based on Comprehension: it would only imply the existence of properties corresponding to formulas.

(Despite its similarity to Naive Comprehension, this principle does not lead to Russell's paradox since: "$\sim F(F)$" is not a grammatical formula of second-order logic and hence isn't substitutable for ϕ; also, substitutions for ϕ cannot have free occurrences of F.) An instance of Comprehension is:

$$\exists F \forall x (F(x) \leftrightarrow x \in B)$$

which, together with (2), implies (1).

Thus in the second argument we still needed a schema, namely the logical axiom schema of comprehension, to reach the desired conclusion. Now, as we saw, the presence of this schema does not compromise the simplicity of second-order ZF. But there is a question of strength. Although second-order ZF has simple and strong laws governing the existence of *sets*, it would seem to lack simple and strong laws governing the existence of *properties*. Better: its theory of the existence of properties is on a par with first-order ZF's theory of the existence of sets. It is schematically simple; and it is somewhat strong, but only as strong as a schema in an enumerable language can be. The failure of strength occurs, to be sure, within the logical part of the theory (if we count second-order logic as logic); but it's hard to see how that matters.

The challenge facing the ZF argument, then, is to say why the elimination of the weakness in first-order ZF justifies the posit of the second-order quantifiers, when the resulting theory, second-order ZF, is weak in a parallel way.

The second-orderist might try to meet the challenge by saying that, while there is a way to eliminate the weakness in first-order ZF (by moving to the second order), there is no conceivable way to eliminate the weakness in second-order ZF. But that just isn't true. We could, for instance, posit a sort of "super-second-order quantification", and use it to state a principle of plenitude for the original second-order quantifiers: for any super-property there is a corresponding property. The second-orderist will presumably regard such additions as misguided, since they too will need a comprehension schema, just as the second-order quantifiers needed one, and thus will have a weakness structurally like the one they were trying to avoid in first-order ZF. But this invites the question of whether we should say the same thing about the shift from first- to second-order ZF and the addition of the second-order quantifiers—that *this* addition is also misguided, given the structural similarity between the weakness of the Comprehension schema in second-order logic and the weakness of the Separation schema in first-order ZF that we were trying to avoid. Why add the second-order quantifiers to deal with a problem which will only re-arise?

It's important to be clear about the structure of the dialectic. I am not denying that second-order ZF would be a stronger theory (in the sense of "strength" relevant to theory choice as described above), if its vocabulary were in good standing. It would indeed be stronger, for it would tell us that there is a set for every property, whereas first-order ZF does not tell us this—it cannot since quantification over properties is not expressible in this language.[23] But notice that the description of the added strength uses the very vocabulary (namely, second-order quantification) whose legitimacy is at issue. It isn't as if the first-orderist can be convicted by his own lights of leaving some lawlike generalization out of his theory, since the allegedly omitted generalization ("there is a set for every property") isn't even statable in his language.

The second-orderist claims that first-order ZF has a kind of "gap" (because certain of its existence claims about sets must be merely schematic) and that this gap should be filled by this law-like statement: "for every property, F, there is a set, z, such that an object x is a member of z iff $F(x)$". The gap-filler is not stateable in the vocabulary of first-order ZF, and thus its existence as a possible content is not common ground; the second-orderist is trying to simultaneously persuade us of the existence of the gap and the means to fill it. But second-orderists are unmoved by the attempt to simultaneously persuade them of the existence of a gap in their theory and the means to fill it: a missing law-like statement "for any super-property $\mathscr{F}$, there is an F such that for all x, $F(x)$ iff $\mathscr{F}(x)$", and the associated additional vocabulary of super-second-order quantification. The crucial point is that the allegation of weakness involves allegedly omitted content that is only recognized by the accuser.

The second-orderist is holding a carrot on a stick: here is some new vocabulary you could adopt (the second-order quantifiers), and in terms of it, new generalizations you could state. Moreover, an axiom schema you accept (the subset axiom schema) can be subsumed under a single principle statable in the new vocabulary (each instance of the subset schema is entailed by the second-order subset axiom; the argument of course uses the second-order comprehension schema). But the carrot still dangles, even after the move to the second order. When we appreciate this dialectical situation, we should realize that we should never have followed the carrot in the first place. For after all, it's on a stick; it will be forever out of reach.

(Some will insist that the notion of a *property* is intuitively legitimate, or

[23] I am not skeptical of the possibility of the higher-order vocabulary being in good standing; Skolemite metasemantic challenges are not at issue here.

well-understood, or needed, or whatever, in a way that neither the notion of set nor the notion of super-second-order quantification are; and that this breaks the symmetry I've been emphasizing between the move from first- to second-order ZF, and the move from second- to super-second-order ZF. Fair enough, but that would be a different argument. I'm here just replying to the ZF argument taken in isolation.)

It is fruitful to compare this dialectic with a similar one involving laws of nature. Consider the law that like-charged particles repel each other. "Deflationists" about laws of nature, such as defenders of the Humean or best-system theory, think that this amounts to nothing more than the regularity that all like-charged particles in fact repel, plus some bells and whistles.[24] "Inflationists" about laws, on the other hand, think that there is some kind of further fact, the law, which explains the regularity that like-charged particles repel. But even the inflationists reject an extreme inflationism, according to which a good explanation of the regularity requires a still further fact, a **Meta-Law** governing the law; for the **Meta-Law** seems to be explanatorily superfluous. Further, although the deflationists regard the inflationist's robust law as explanatorily superfluous, even they tend to reject the extreme deflationism of someone like Michael Esfeld (2020), who thinks that the posit of charge is superfluous since we could just as well state the law of motion by saying (roughly) that particles move in certain ways, namely the ways they would move if there were a property of charge; according to Esfeld, adding that the differences in motion are due to differences in charge does not improve the explanation. It is not obvious who is right in this dialectic; the answer turns on difficult questions about explanation.[25] But I suspect that most will agree either with the standard deflationist or the standard inflationist, and will reject both extreme deflationism (on the grounds that eliminating charge from physics is a genuine explanatory loss) and certainly extreme inflationism (on the grounds that the **Meta-Law** is explanatorily superfluous).

To my mind, the comparison to this dialectic about laws of nature weakens the ZF argument. What seems particularly objectionable about extreme inflationism is the parallelism between its explanation of regularities and the inflationist's explanation: the two are exactly alike except for an added layer in the former case. But the same objection, as we saw, seems to apply to the defender of second-order ZF. Unlike the positing of charge, which *does* result

[24] See …

[25] See Dorr (2007, section 3; 2010, p. 160–3) and Sider (2020, section 4.12).

in a structurally different explanation (this is partly why Esfeld does not convince), the positing of the second-order quantifiers does not seem to result in a structurally different explanation.[26]

5. The argument from semantics

The status of the ZF argument, in my view, is parallel to the status of another argument for higher-order languages, namely that such languages are needed to give a semantic account of languages with unrestricted quantifiers.[27]

The standard approach to semantics in logic—to giving an account of how sentences come to be true and false, and of when sentences imply one another in virtue of meaning—is the model-theoretic approach. A model is normally defined as an ordered pair $\langle D, F \rangle$, where D, the domain, is a set, and where F, the interpretation function, is a function that assigns to each nonlogical expression in the language some appropriate set-theoretic construction based on D: members of D to names, and sets of n-tuples of D (extensions) to predicates. Using methods developed by Tarski, one can define what it means for an arbitrary sentence of the language to be true in such a model.

Thus the standard approach defines the domain of any model, and the extension of any predicate in any model, as sets. But that is limiting. In the intended interpretation of the language of first-order ZF set theory, for instance, the quantifiers range over all sets and the predicate '$\in$' stands for set membership. So the domain of a model corresponding to this intended interpretation should be a set containing all sets, and the extension of '$\in$' in this model should be a set containing all and only ordered pairs $\langle y_1, y_2 \rangle$ where y_1 is a member of y_2. But there are no such sets. ZF set theory (which is assumed in the metalanguage) says that no such sets exist (for the assumption that either set exists would lead to Russell's contradiction).

Given the standard approach, then, no model captures the intended interpretation of the language of first-order set theory. At best, models can provide a distorted representation of that intended interpretation, by treating the quantifiers as being restricted to a mere part of the set-theoretic hierarchy.[28]

[26] To be sure, it would be nice to have a clear account of "structurally different explanation".

[27] See, for instance, Williamson (2003).

[28] There are subtle arguments that this limitation does not affect which sentences the standard approach counts as valid or imply one another. But such arguments are less clearly correct when the languages move beyond the first order, and break down if the language contains

Against this backdrop, the higher-order outlook becomes attractive. For one can, in a second-order language, define a sort of "model" in which the "domain" can contain all the sets, and in which the "extension" of '$\in$' contains all and only the ordered pairs $\langle y_1, y_2\rangle$ where y_1 is a member of y_2. The trick is to abandon first-order quantification over models conceived as entities, and instead to treat quantification over models as being second-order.[29] We define "R is a model", for R a second-order dyadic variable, in such a way that $R(x, y)$ means that y is a semantic value of the linguistic expression x. So if 'S' is a two-place predicate, then $R(\text{`}S\text{'}, \langle y_1, y_2\rangle)$ can be thought of as meaning that $\langle y_1, y_2\rangle$ is "in the extension of 'S' in R", although this is misleading since we are not accepting the existence of an entity, the extension of 'S' in R. In one of these second-order "models" the "extension" of the two-place predicate '$\in$' will consist of all and only the ordered pairs whose first coordinate is a member of the second coordinate. That is: for some R of the relevant sort, $R(\text{`}\in\text{'}, y)$ if and only if y is an ordered pair $\langle y_1, y_2\rangle$ such that y_1 is a member of y_2. No paradox results because we do not recognize an entity as the extension of '$\in$'. Williamson then shows how to characterize a notion of a sentence being true in such a "model" R.

By adopting a second-order language, we can thereby state more adequate semantic theories for first-order languages. But we will then naturally aspire to state semantic theories for second-order languages, such as the one we use to state semantic theories for first-order languages. And as Øystein Linnebo and Agustín Rayo (2012) argue, if we want to acknowledge the full range of "models" for second-order languages, a third-order metalanguage will be needed. Let α be some one-place second-order predicate constant in the second-order language (i.e., a predicate that can be attached to one-place predicates—i.e., type $((e, t), t)$). For any property, $\mathscr{F}$, of properties (variable of type $((e, t), t)$), it would be possible to interpret α so that it applies to exactly the properties G (variable of type (e, t)) such that $\mathscr{F}(G)$. So for each such $\mathscr{F}$, there must be a new model R. But it can be shown that there are more such $\mathscr{F}$s than there are second-order relations R. (The argument is analogous to the usual Cantorian diagonal argument showing that a set has strictly lower cardinality than its power set, but here the "cardinality comparison" and argument are made in a higher-order language.) The models for the second-order language must be

certain sorts of expressions. And even if it always gives the right answers, the standard approach would seem still to be mis-modeling the semantic facts. See Boolos (1985); McGee (1992); Rayo and Uzquiano (1999) for discussion.

[29] See Boolos (1985); Rayo and Uzquiano (1999); Williamson (2003).

third-order relations.[30]

In fact, Linnebo and Rayo show how to generalize this argument into the transfinite. For each language in a certain transfinite hierarchy (and thus languages with syntactic types beyond those that I defined earlier, which were all finite), stating its semantics requires a still higher order metalanguage.

This fact undermines the original argument for the higher-order viewpoint, if that argument is taken in the metaphysical spirit we have been considering. We began with an explanatory ambition: to give a certain sort of semantic theory for the language of first-order set theory. If we take that ambition to justify recognizing second-order quantification, with its attendant worldly complexity, then we are saddled with a new explanatory ambition, the satisfaction of which requires a new language, which results in a new explanatory ambition; and so on. The explanatory demand is insatiable, in that there is no language we could speak in Linnebo and Rayo's hieararchy in which we could state a semantics for all languages in that hierarchy.

Eventually we will need to resist the explanatory demand. At that point, we will have recognized all the worldly complexity of the preceding higher-order languages, but will be speaking some final language L_α (for some ordinal α) for which we can give no semantics. If we must live with this situation for L_α eventually, wouldn't it have been better to embrace it from the start, with the first-order language of set theory? Embracing the languages up to V_α seems akin to embracing, in addition to robust laws, also meta-laws, meta-meta-laws, and so on, up to a certain point and then stopping; or embracing, in addition to sets, second-order quantification, super-second-order quantification, super-duper-second-order quantification, and so on, and then stopping. The added value of the additional explanations in each case is dubious.

Since a semantic theory of the kind we were seeking is ultimately going to be unavailable for some language we speak, we should live with this from the start, with the language of first-order set theory. This doesn't mean that that language is meaningless, or doesn't quantify over all sets after all; it rather

[30] The argument, notice, is that a higher-order metalanguage is needed to recognize all the possible interpretations for the original language. If we only wished to give the *intended* interpretation of the second-level language, a second-level metalanguage could be used (although given Tarski's theorem it would require a new primitive semantic predicate). But an explanatory theory of meaning should include a theory of all the semantic possibilities, of how variation in the meanings of the parts induces variation in the meanings of wholes, just as an explanatory physical theory should encompass all the physical possibilities, not just the actual one. Good explanations must be general.

means that a certain sort of *theory* of the meaningfulness of that language is unavailable. Nor does it mean that no theory of its meaningfulness is available: the usual models of model theory can still be models in the philosophy-of-science sense, albeit imperfect ones, of meaning. It's a bit sad, but not the end of the world, and in any case ultimately unavoidable.

6. Higher-order definition of necessity

One idea that has emerged in the higher-order literature is that the higher-order quantifiers could be used to give a reductive definition of necessity. Let $\top$ be some arbitrarily chosen logical truth, $\forall x\, x = x$, say. $\Box A$ is then defined as meaning that $A = \top$. Necessity is identity-to-$\top$.[31]

The availability of such a definition might be taken to be a point in favor of the higher-order framework. The argument would be parallel to the common argument on behalf of sets: given set theory we can define the rest of mathematics.[32]

If we are to identify necessity with identity-to-$\top$, then the latter must obey any logical principles obeyed by the former. For instance, it is usually assumed that first-order logical truths are necessary, and that necessity obeys the "K" principle that $\Box(A \to B) \to (\Box A \to \Box B)$. Thus first-order logical truths must be identical to $\top$; and it must be that if $(A \to B)$ and A are each identical to $\top$, then so is B. Each of these assumptions follows from a general thesis that Andrew Bacon and Dorr call "Classicism", which implies that $A = B$ whenever $A \leftrightarrow B$ is a first-order logical truth.[33] Let's build Classicism into this account of necessity.

Certain sorts of necessity are clearly distinct from identity-to $\top$. Certainly "deontic necessity" is not identity-to-$\top$: some deontically necessary statements are false, whereas anything that is identical to $\top$ is true. Even amongst the

[31]That is: $\Box = \lambda p.(p =^t \top)$.

[32]Higher-order definitions of other metaphysical concepts have been proposed, such as fundamentality and priority; the availability of such definitions might be regarded as offering similar support to the higher-order framework.

[33]More accurately, the assumptions follow given a certain background higher-order logic, which includes the assumption that $=$ (i.e., $=^t$) obeys Leibniz's Law. Suppose A is a first-order logical truth. Then $A \leftrightarrow \top$ is also a first-order logical truth, and so $A = \top$, given Classicism. Next, suppose that $(A \to B) = \top$ and $A = \top$. Then by Leibniz's Law, $(\top \to B) = \top$. But $(\top \to B) \leftrightarrow B$ is a first-order logical truth; so by Classicism, $(\top \to B) = B$; and so, by Leibniz's Law, $B = \top$.

alethic (truth-requiring) modalities, such as technological necessity, physical necessity, metaphysical necessity, etc., they can't all be identified with identity-to-$\top$ since they would then be identified with each other. It is *metaphysical* necessity that is identified with identity-to-$\top$.

The identification must be taken in the right spirit. The higher-orderist's attitude is *not* that it will be obvious to everyone that identity-to-$\top$ is what we have meant all along by 'metaphysically necessary', or even that its extension mostly fits standard assumptions about 'metaphysically necessary'. The attitude is rather that the usual notion of metaphysical necessity is not at all in good standing, that it is very unclear what 'metaphysical necessity' is supposed to mean, and that identity-to-$\top$ is the closest clear notion. We should mean identity-to-$\top$ by 'metaphysically necessary' on pain of meaning nothing (or nothing determinate) at all.[34]

What is usually said to introduce the notion of metaphysical necessity is indeed thin. Saul Kripke (1972, p. 99) glosses it as meaning "necessary in the highest degree—whatever that means"; Alvin Plantinga (1974, p. 1) says "we must give examples and hope for the best". Then again, when asked to single out *any* of the central concepts of philosophical importance, the result is invariably thin. Many ethicists think that there is an important notion of being morally obligatory, but it is hard to say (without taking stands on controversial questions) exactly which notion of obligation, out of all the possible notions of obligation—legal, familial, pertaining to etiquette, etc., let alone other "sorts of obligation" for which we have no names—the notion of *moral* obligation is supposed to be. Nor is it easy to say what a "notion of obligation" is. Rather than replacing the notion of moral obligation with a radically different but logically hygienic notion, one might instead live with the fact that this important notion isn't easy to single out in noncircular or enlightening terms.[35]

[34]Some higher-orderists have a similar attitude about a number of other metaphysical concepts, such as fundamentality, ground, and priority. "The transition to the new regime may not be an easy one for some of you, but it is for your own good." There is a bit of French's (2014) "Viking" approach at work here.

[35]The concern about 'metaphysically necessary' can be given a metasemantic spin: given how little we can say to single out its intended meaning, won't it either refer to nothing or else be hopelessly indeterminate? But the defender of metaphysical necessity could reasonably invoke "reference magnetism" in response: considerations like "fit with how we use the term" are only part of the story of how meaning gets determined; another part of the story is that the world itself plays an important role. If there is a candidate meaning that "carves at the joints", then an expression can mean that candidate meaning even if our usage of the expression doesn't favor that candidate as opposed to various other candidates. See Lewis (1984, 1983*a*),

For that matter, the quantifiers themselves (whether first- or higher-order) are in a similar boat. We can say a bit about their logical behavior, for instance that existentials follow from their instances. But even that isn't as helpful as it might seem since sentences containing names like 'Santa Claus' arguably create problems for the rule of existential introduction, but arguably seem to be disqualifiable only in quantificational terms: 'Santa Claus' doesn't refer to anything. The usual rules of inference fall far short of securing the meanings for quantifiers in at least this sense: when faced with a difficult ontological question, such as whether there exist gods, or sets, or medium-sized dry goods, being told the inference rules is next to useless.[36] I myself share the sense that the first-order quantifiers, anyway, are firmer ground on which to base philosophy than 'metaphysical necessity' (they are after all central to pretty much all inquiry). Maybe the higher-order quantifiers are likewise firm ground, but this is a bold hypothesis and an open question, not settled by the thinness of the way rival ideology is introduced.

How much does identity-to-$\top$ fit our ordinary beliefs about metaphysical necessity?

If necessity is identity-to-$\top$ then all necessary propositions are identical, which might seem absurd. It is necessary that $2+2=4$, and it is necessary that all scarlet things are red, but aren't $2+2=4$ and All scarlet things are red distinct propositions?

In fact, Classicism undermines the most natural reasons one might give for thinking that these propositions are distinct. For instance, one might be inclined to say that the property of redness is a *part* of the proposition that all scarlet things are red, but not part of the proposition that $2+2=4$; or that the first but not the second proposition is *about* redness. But these judgments about parthood and aboutness become problematic once one realizes that, for instance, $(p\vee{\sim}p)=(q\vee{\sim}q)$, for any p and q, given Classicism. Thus the proposition that snow is white or not white is identical to the proposition that

---

and also Williams (2007); Sider (2011, section 3.2). I wouldn't myself defend the determinacy of 'metaphysically necessary' in this way; see Sider (2011, chapter 12)

[36]It may be argued that the "collapse arguments" are relevant here. These arguments purport to show that any two candidate meanings for the quantifiers are in fact equivalent. I criticized these arguments in Sider (2007, pp. 217–18), but Dorr (2014) has pushed the subject significantly further. I hope to discuss this in the future. I will say this: Dorr's best version of the collapse argument in the case of the first-order quantifiers uses higher-order quantifiers in a way that I suspect limits the argument's value when applied to higher-order quantifiers generally.

roses are red or not red; and so the fact that one sentence contains the word 'red' whereas another does not is not sufficient reason to say that the proposition expressed by the first is "about" redness, or "contains redness as a part", whereas the proposition expressed by the second does not. Indeed, one might conclude that these notions of aboutness and propositional parthood do not even make sense, given Classicism. Of course, one might go the other way and reject Classicism because of examples like those we are considering, and for that reason reject the identification of necessity with identity-to-$\top$. But Classicism is a simple and nonarbitrary and somewhat intuitive idea about propositional grain. Moreover, the most obvious way of making the notions of aboutness and propositional parthood precise, namely a certain "structured" conception of propositions (and other higher-order "entities"), is self-contradictory given the higher-orderists' usual assumptions about higher-order logic: this is the "Russell-Myhill paradox".[37]

In addition to this powerful strategy for defending against objections to the surprising identifications, there is also an offensive strategy available in certain cases. It is, let us suppose, metaphysically necessary that all gold has protons. Thus if necessity is identity-to-$\top$, the proposition that all gold has protons must be identical to $\top$. Though this identification is surprising, in fact one can argue for it, given Classicism. There is good evidence that being gold is identical to having atomic number 79, and that having atomic number 79 is identical to having seventy nine protons. These are property identities, understood in higher-order terms (thus regimented with λ and $=^{\langle e,t\rangle}$, the identity predicate that can be flanked by one-place first-order predicates), and may be supported by familiar arguments from Kripke (1972) and David Lewis (1966). Now, this is a first-order logical truth:[38]

Everything that has seventy nine protons has protons

So this is also a first-order logical truth:

Everything that has seventy nine protons has protons $\leftrightarrow \top$

So given Classicism:

Everything that has seventy nine protons has protons $= \top$

[37] See Fritz (2017) for a nice presentation.

[38] I am understanding "y has 79 protons" in its first-order logic, nonmathematical sense: $\exists x_1 \ldots \exists x_{79}\big(\mathrm{Proton}(x_1) \,\&\, \cdots \,\&\, \mathrm{Proton}(x_{79}) \,\&\, x_1 \neq x_2 \,\&\, \cdots \,\&\, \mathrm{has}(y, x_1) \,\&\, \cdots \,\&\, \mathrm{has}(y, x_{79})\big)$.

But the property identities mentioned earlier imply this property identity:

has seventy nine protons $=$ is gold

Then by Leibniz's Law, the previous two statements imply:

Everything that is gold has protons $= \top$

But in many cases of statements traditionally thought to be necessary, this offensive strategy isn't available. Following Dorr (2005, p. 263), call a sentence *metaphysically analytic* if it can be transformed into a first-order logical truth by substituting an expression α of any type for another expression β of that type, where $\ulcorner \alpha = \beta \urcorner$ is true. The offensive strategy amounts to showing that a proposition is necessary by showing that it is metaphysically analytic, via some reasonably uncontroversial identity. Metaphysical analyticity is a metaphysical analog of the old idea of analyticity, understood as transformability into a logical truth by substitution of synonyms. Now, the idea that necessity is analyticity was traditionally thought to be threatened by various putative cases of synthetic necessity. Many such cases appear also to be metaphysically synthetic (or rather, cannot uncontroversially be shown to be metaphysically analytic), and thus resist the offensive strategy.

For instance, consider the old chestnut "nothing is both red and green". (That is: nothing is both uniformly red all over its surface and uniformly green all over its surface.) It does not seem possible to transform this sentence into a logical truth by substituting using identities. The only possible route would seem to be substituting via the putative identity 'green $=$ green and not red' in the logical truth 'Nothing is both red and green and not red'. But it isn't the case that to be green is to be green and not red, at least not clearly so.

Another example: "For all x, y, and z, if x is part of y and y is part of z then x is part of z". Again, it's hard to find an identity via which this is metaphysically analytic. The best bet would seem to be if parthood is identical to the ancestral of some relation, R—that is, if the following is true for some R

$$\text{is part of } =^{\langle e,t\rangle} \lambda xy.\forall F\big((\forall z(Rzy \to Fz) \,\&\, \forall z \forall w((Fw \,\&\, Rzw) \to Fz)) \to Fx\big)$$

But what might the relation R be? One possibility is that it is some relation of *immediate* parthood. But perhaps there is no such relation. Another possibility is that R is parthood itself; but then the alleged identity feels off, just as 'green $=$ green and not red' did.

A third example: assume a "platonist" conception of mathematical objects, and consider either the (first-order) continuum hypothesis or its negation, whichever is true. It's hard to see how this sentence could be shown via some relatively uncontroversial identity to be metaphysically analytic.

Finally, consider true statements of metaphysics, such as 'For every x and y, there exists a mereological sum of x and y' (or, if you like, 'nothing is a proper part of anything'); or 'for all x, x exists now' (or, if you like, its negation). These and related statements also seem resistant to being shown to be metaphysically analytic.[39]

Now, the dialectic here is complex. For the defensive strategy mentioned above, based on Classicism, can be used to defend identities that would show that the sentences I've been discussing are metaphysically analytic after all. For instance, the Classicist might ask: what makes you so sure that green is not identical to green and not red? Reasoning based on notions like aboutness or propositional parthood is undermined by Classicism, as we saw.

But the question at the moment is how far the *offensive* strategy can take us. The offensive strategy aims to convince a neutral party of surprising identifications—identifications that we don't find initially compelling or intuitive—using Classicism and identifications that *do* seem initially compelling. But the ancillary identifications needed to complete the arguments—identifications such as that green = green and not red—are not initially compelling. And the fact that those ancillary identifications can be defended against certain objections by means of Classicism does not make them compelling, i.e., useable in a dialectically effective offensive argument.

It's perhaps also worth emphasizing that although one forgoes certain natural arguments for distinctness (namely, the ones that run through notions of aboutness or propositional parthood) when one embraces Classicism, that doesn't mean that just anything goes, that one should be open to just any identifications that aren't ruled out by the rest of one's theory. It may remain reasonable to hold on to one's initial conviction that green is distinct from green and not red, that neither the continuum hypothesis nor its negation is identical to $\forall x\, x = x$, and so on, even in the absence of principled arguments against these identifications.

But let's end this somewhat tiresome discussion of dialectic and burden of proof, and turn to a final, more big-picture observation. There is a traditional

[39]Though Dorr (2005) provides an interesting argument for the metaphysical analyticity of 'nothing is part of anything'.

concern about reductive theories of necessity, which is that they tend to leave out genuine "mustness".[40] (The concern is akin to equally traditional concerns about reductive conceptions of morality and laws of nature.) This concern extends even to logical truths: they seem to be necessary in some robust and nontrivial sense. But if necessity is identity-to-$\top$, they are necessary only in the trivial sense in which they are identical to themselves. Put another way, the identification of necessity with identity-to-$\top$ presupposes, rather than accounting for, the genuine necessity of logical truths. I myself am not sympathetic to this sort of argument, but I suspect that many readers will be.

References

Almog, Joseph (1989). "Logic and the World." In Joseph Almog, John Perry and Howard Wettstein (eds.), *Themes from Kaplan*, 43–65. New York: Oxford University Press.

Blackburn, Simon (1987). "Morals and Modals." In *Fact, Science and Morality: Essays on A. J. Ayer's Language, Truth and Logic*. Oxford: Blackwell. Reprinted in Blackburn 1993.

— (1993). *Essays in Quasi-Realism*. Oxford: Oxford University Press.

Boolos, George (1975). "On Second-Order Logic." *Journal of Philosophy* 72: 509–27. Reprinted in Boolos 1998: 37–53.

— (1984). "To Be Is to Be the Value of a Variable (or to Be Some Values of Some Variables)." *Journal of Philosophy* 81: 430–49. Reprinted in Boolos 1998: 54–72.

— (1985). "Nominalist Platonism." *Philosophical Review* 94: 327–44. Reprinted in Boolos 1998: 73–87.

— (1998). *Logic, Logic, and Logic*. Cambridge, MA: Harvard University Press.

Carroll, Lewis (1895). "What The Tortoise Said To Achilles." *Mind* 4: 278–80.

Church, Alonzo (1940). "A Formulation of the Simple Theory of Types." *Journal of Symbolic Logic* 5: 56–68.

[40] See Blackburn (1987).

Dorr, Cian (2005). "What We Disagree About When We Disagree About Ontology." In Mark Eli Kalderon (ed.), *Fictionalism in Metaphysics*, 234–86. Oxford: Oxford University Press.

— (2007). "There Are No Abstract Objects." In Theodore Sider, John Hawthorne and Dean W. Zimmerman (eds.), *Contemporary Debates in Metaphysics*, 32–63. Oxford: Blackwell.

— (2010). "Of Numbers and Electrons." *Proceedings of the Aristotelian Society* 110(2pt2): 133–81.

— (2014). "Quantifier Variance and the Collapse Theorems." *The Monist* 97: 503–70.

— (2016). "To Be F is to Be G." *Philosophical Perspectives* 30(1): 39–134.

Esfeld, Michael (2020). "Super-Humeanism: The Canberra Plan for Physics." In David Glick, George Darby and Anna Marmodoro (eds.), *The Foundation of Reality: Fundamentality, Space and Time*, 125–38. Oxford: Oxford University Press.

Fine, Kit (2001). "The Question of Realism." *Philosophers' Imprint* 1: 1–30.

French, Steven (2014). *The Structure of the World: Metaphysics and Representation*. Oxford: Oxford University Press.

Fritz, Peter (2017). "How Fine-Grained is Reality?" *Filosofisk Supplement* 13(2): 52–7.

Gabriel, Gottfried, Hans Hermes, Friedrich Kambartel, Christian Thiel, Albert Veraart, Brian McGuinness and Hans Kaal (1980). *Gottlob Frege: Philosophical and Mathematical Correspondence*. Oxford: Blackwell.

Grover, Dorothy (1992). *A Prosentential Theory of Truth*. Princeton: Princeton University Press.

Hilbert, David (1899). "Letter to Frege." In Gabriel et al. (1980).

Hirsch, Eli (2011). *Quantifier Variance and Realism: Essays in Metaontology*. New York: Oxford University Press.

Kripke, Saul (1972). "Naming and Necessity." In Donald Davidson and Gilbert Harman (eds.), *Semantics of Natural Language*, 253–355, 763–9. Dordrecht: D. Reidel. Revised edition published in 1980 as *Naming and Necessity* (Harvard University Press, Cambridge, MA).

Lewis, David (1983*a*). "New Work for a Theory of Universals." *Australasian Journal of Philosophy* 61: 343–77. Reprinted in Lewis 1999: 8–55.

— (1983*b*). *Philosophical Papers, Volume 1*. Oxford: Oxford University Press.

— (1984). "Putnam's Paradox." *Australasian Journal of Philosophy* 62: 221–36. Reprinted in Lewis 1999: 56–77.

— (1999). *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.

Lewis, David K. (1966). "An Argument for the Identity Theory." *Journal of Philosophy* 63: 17–25. Reprinted in Lewis 1983*b*: 99–107.

Linnebo, Øystein and Agustín Rayo (2012). "Hierarchies Ontological and Ideological." *Mind* 121(482): 269–308.

McGee, Vann (1992). "Two Problems with Tarski's Theory of Consequence." *Proceedings of the Aristotelian Society* 92(n/a): 273–92.

McSweeney, Michaela Markham (2019). "Logical Realism and the Metaphysics of Logic." *Philosophy Compass* 14(1): e12563.

Paul, L. A. (2012). "Metaphysics as Modeling: The Handmaiden's Tale." *Philosophical Studies* 160: 1–29.

Plantinga, Alvin (1974). *The Nature of Necessity*. Oxford: Oxford University Press.

Prior, A. N. (1971). "Platonism and Quantification." In *Objects of Thought*, 31–47. Oxford: Oxford University Press.

Quine, W. V. O. (1948). "On What There Is." *Review of Metaphysics* 2: 21–38. Reprinted in Quine 1953: 1–19.

— (1953). *From a Logical Point of View*. Cambridge, MA: Harvard University Press.

— (1970). *Philosophy of Logic*. Cambridge, MA: Harvard University Press. Second edition, 1986.

Rayo, Agustín and Gabriel Uzquiano (1999). "Toward a Theory of Second-Order Consequence." *Notre Dame Journal of Formal Logic* 40: 315–25.

Rayo, Agustín and Stephen Yablo (2001). "Nominalism through De-Nominalization." *Noûs* 35: 74–92.

Sider, Theodore (2007). "Neo-Fregeanism and Quantifier Variance." *Aristotelian Society, Supplementary Volume* 81: 201–32.

— (2011). *Writing the Book of the World*. Oxford: Clarendon Press.

— (2020). *The Tools of Metaphysics and the Metaphysics of Science*. Oxford: Oxford University Press.

Thomasson, Amie L. (2007). *Ordinary Objects*. New York: Oxford University Press.

— (2015). *Ontology Made Easy*. New York: Oxford University Press.

Turner, Jason (2015). "What's So Bad About Second-Order Logic?" In *Quantifiers, Quantifiers, and Quantifiers. Themes in Logic, Metaphysics, and Language*, 463–87. Springer.

Whitehead, Alfred North and Bertrand Russell (1910). *Principia Mathematica*. Second edition. Cambridge: Cambridge University Press, 1925.

Williams, J. Robert G. (2007). "Eligibility and Inscrutability." *Philosophical Review* 116: 361–99.

Williamson, Timothy (2003). "Everything." *Philosophical Perspectives* 17: 415–65.

— (2013). *Modal Logic as Metaphysics*. Oxford: Oxford University Press.