Composition as Abstraction

Jeffrey Sanford Russell

August 2016 draft

I'll present an argument for Mereological Universalism: the thesis that any things whatsoever have a mereological sum (or *fusion*). One reason people present arguments is to try to convince people of the truth of their conclusions. That's not what I'm up to. Instead, the point is to bring out a connection between one interesting claim and another: in this case, to explore a connection between the debate over composition and another debate in the philosophy of mathematics. I'll begin by going over some suggestive technical facts, and discuss their philosophical significance in Section 3.

1 Mereology of the Many

The Los Angeles metropolitan area (as statistically defined) is larger than just the city of Los Angeles, and it is larger than Los Angeles County. Parts of it extend into five different counties (Los Angeles, Ventura, Orange, Riverside, and San Bernardino). Besides the cities and suburbs that make up the LA metro area, those five counties also include large sparsely populated swaths of farmland, desert, and mountain ranges. The LA metro area isn't exactly *composed* of the five counties: taken together, they are larger than it. But it is *covered* by them. The metro area includes more than a hundred incorporated cities (including the city of Los Angeles, Long Beach, Glendale, ...). These cities are *collectively* covered by the five counties. Covering, in this sense, is a plural-plural relation. It is a plural analogue of having parts, done by many things collectively rather than by one thing all on its own.

Let's consider how we can define this plural relation more precisely—and in a way that does not build in assumptions about the existence and structure of parts and wholes. The background is standard plural logic (see e.g. Linnebo 2014).

Start with an "ur-part" relation. We'll suppose is transitive and reflexive, but we make no other assumptions about its structure. We can define a corresponding *plural* relation in terms of ur-parthood:

Thanks to Tom Donaldson, Jeremy Goodman, Gabriel Uzquiano, and two anonymous referees for comments.

> The Y's **cover** the X's ($X \leq Y$) iff anything that shares an ur-part with some X also shares an ur-part with some Y.

To get the idea, note that if ur-part satisfies classical mereology, then $X \leq Y$ iff the fusion of the X's is part of the fusion of the Y's. But the definition of covering clearly makes no appeal to the existence of fusions. At the opposite extreme, suppose ur-part is the relation of identity. In that case, $X \leq Y$ iff each of the X's is one of the Y's.

This definition has some nice consequences. For any X's, Y's, and Z's:

> **Reflexivity.** $X \leq X$.
>
> **Transitivity.** If $X \leq Y$ and $Y \leq Z$ then $X \leq Z$.

We can also define:

> The X's **overlap** the Y's iff there are some Z's such that $Z \leq X$ and $Z \leq Y$.

Then we can conclude furthermore:

> **Supplementation.** If for any Z's that overlap the X's, the Z's overlap the Y's, then $X \leq Y$.
>
> **Fusions.** If there are some X's such that $\varphi(X)$, then there are Y's such that, for any Z's, the Z's overlap the Y's iff the Z's overlap some X's such that $\varphi(X)$. (In short, if there are any ϕ-pluralities, some plurality *fuses* the ϕ-pluralities.)

Fusions is a schematic principle: $\varphi(X)$ can be filled in with any formula of one plural variable.[1]

For those who have spent much time pondering part and whole, this is suggestive: what we have stated are precisely an axiomatization of classical mereology—without anti-symmetry, to which I'll return—the key difference being that we have used *plural* variables everywhere instead of *singular* variables. This kind of "plural mereology"

[1] The proofs of these facts are left as an exercise. Here are two useful intermediate lemmas:

1. If x is an ur-part of some Y, then $\{x\} \leq Y$. (Here the curly brace notation stands for the "singleton-plurality" of just those things which are x.)
2. The X's overlap the Y's iff some X and some Y share an ur-part.

For the proof of Fusions, we can let the Y's be just those things x such that, for some X's, x is one of the X's and $\varphi(X)$ (using plural comprehension).

The technical idea here is similar to a construction presented by Paul Hovda (2014; 2016).

is as innocent as plural quantification. Whatever parthood may be, there is a natural extension of it to pluralities which obeys classical (but perhaps non-extensional) mereology.

Classical mereology minus anti-symmetry—that is, the theory axiomatized by the singular versions of Reflexivity, Transitivity, Supplementation, and Fusions—doesn't seem to have a standard name. I'll call it **Non-Extensional Mereology**. This contrasts with **classical (extensional) mereology**, which adds the further axiom that parthood is anti-symmetric. (The name is not perfect: note that Non-Extensional Mereology is *compatible* with extensionality, though it does not imply it. Parallel naming issues also arise for other mathematical terms: for example, non-commutative geometry is compatible with commutativity, and total functions are a special case of partial functions.) (For discussion of the relationship between extensionality and anti-symmetry, see Cotnoir 2010.)

This is nice, but it shouldn't be terribly surprising. Indeed, opponents of unrestricted composition standardly appeal to plural relations as a surrogate for relations among sums. For instance, Peter van Inwagen paraphrases "Some chairs are heavier than some tables" as "There are *x*s that are arranged chairwise and there are *y*s that are arranged tablewise and the *x*s are heavier than the *y*s," where in the paraphrase, "heavier" means *collectively* heavier—a certain plural relation (Inwagen 1994, 109). The cogency of "plural mereology" shows that this general paraphrase strategy can be applied to a theory of mereology, in particular. In the case of parthood, unlike the case of heaviness, we get the nice extra result that the plural paraphrases of the principles of Non-Extensional Mereology turn out to be truths of plural logic.

The debate over composition centrally relies on a distinction between there being some things collectively behaving a certain way and there being *something* individually behaving a certain way. This distinction seems to be in good order. So the friend of arbitrary fusions needs to bridge the gap from "plural mereology" to standard singular mereology.

2 Out of Many, One

By way of analogy, I'll now rehearse a story from the philosophy of mathematics.[2]

[2] There is one important respect in which the version of the story I tell is non-standard. The original context in which Frege framed his proposals about the foundations of mathematics was not plural logic, but rather relational higher-order logic, and the neo-Fregean literature has followed him. I have translated the story into the idiom of plural logic, because this is more familiar and less controversial for mereologists. One side-effect of the translation is that plurally defining one-to-one correspondences

Plural logic gives us resources for describing facts about cardinality. We can define "There are no more X's than Y's" to mean that there is some way of pairing every X with some unique Y. I'll use the notation $X \preceq Y$ for this plural relation (to distinguish it from covering).

> $X \preceq Y$ iff there are some ordered pairs, the Z's, such that for each x among the X's there is exactly one y among the Y's such that (x, y) is a Z-pair, and for each y among the Y's there is at most one x among the X's such that (x, y) is a Z-pair.

We can go on to derive important features of this plural relation, as a matter of pure plural logic.[3] For instance (for any X's, Y's, and Z's):

> **Reflexivity.** $X \preceq X$.
>
> **Transitivity.** If $X \preceq Y$ and $Y \preceq Z$ then $X \preceq Z$.
>
> **Totality.** Either $X \preceq Y$ or $Y \preceq X$.
>
> **Well-Ordering.** If there are some X's such that $\varphi(X)$, then there are X's such that $\varphi(X)$ and for any Y's such that $\varphi(Y)$, $X \preceq Y$. (In short, if there are any φ-pluralities, there is a *smallest* φ-plurality.)

We can also similarly define "plural addition" and other arithmetic relations. (Furthermore, we can define "there are finitely many X's" using Frege's definition of the ancestral, and thus by restricting attention to finite pluralities can describe a theory of *finite* plural arithmetic—the usual domain of number theory—rather than the more general infinitary cardinal arithmetic we get most straightforwardly.)

We have derived some standard axioms for a theory of cardinal numbers, modulo capitalization: these logical principles use plural variables rather than the usual sin-

requires appeal to ordered pairs (or something that can play that role). This technical detail, and more generally the differences between plural quantification and relational quantification, only matters for the story about cardinality, not the story about mereology, so I've attempted to keep these details in the background as much as possible. For convenience, I'll go ahead and count the existence of unique ordered pairs of non-pairs as part of "logic". (I won't assume that there are pairs whose elements are themselves pairs; this is just to avoid trivializing Infinity, below.)

[3] This is a bit of an overstatement, since deriving Totality and Well-Ordering depends on strong background principles of plural logic, such as the axiom of Global Well-Ordering, which says that there are ordered pairs that well-order everything. (This is a plural generalization of a standard set-theoretic equivalent of the axiom of Choice. See Linnebo (2010, 161) for discussion and defense of this principle—though this defense presupposes set theory, which in our present context might be tendentious.) We can avoid these technicalities insofar as our concern is with *finite* arithmetic: the restrictions of Totality and Well-Ordering to finite pluralities do not require any such strong axioms.

gular variables. Note also that, as in the mereological case, we cannot derive an Anti-Symmetry axiom.[4]

There are some difficulties for plural arithmetic that don't arise for plural mereology. For instance, standard arithmetic includes the statement "every finite number has a successor", whose plural analogue cannot be proved from pure plural logic. In fact, the plural analogue of this statement is equivalent to

> **Infinity.** There are infinitely many things.[5]

Plural mereology is a bit better off than plural arithmetic, since it does not rely on any such principle.

Aside from the issue of Infinity, there is another way this "plural arithmetic" (which is just a fragment of plural logic) seems to fall short of what we want from a genuine theory of arithmetic. We have a theory of the cardinal structure of pluralities of *things*—but no individual *thing* which is a *cardinal number*. Frege writes:

> [T]he individual number shows itself for what it is, a self-subsistent object. I have already drawn attention above to the fact that we speak of "The number 1", where the definite article serves to class it as an object. In arithmetic this self-subsistence comes out at every turn, as for example in the identity 1 + 1 = 2. … And identities are, of all forms of proposition, the most typical of arithmetic. (Frege 1953, sec. 57)

---

[4]The plural version of Anti-Symmetry would amount to saying that there is just one plurality of any particular size: if $X \preceq Y$ and $Y \preceq X$, then $X = Y$. This is inconsistent with the existence of more than one thing. Suppose $x \neq y$. Then the singleton plurality $\{x\}$ (which consists of just those things which are identical to x) is in one-to-one correspondence with the singleton plurality $\{y\}$. That is, $\{x\} \preceq \{y\}$ and $\{y\} \preceq \{x\}$; but these are distinct pluralities, which contradicts plural Anti-Symmetry.

[5]To justify this equivalence, we begin by defining the plural notion of "successor" (following Frege 1953, sec. 76):

> The Y's *succeed* the X's iff there is some y among the Y's such that the X's are in one-to-one correspondence with those things among the Y's which are distinct from y.

In other words, the X's can be lined up with Y's with exactly one of the Y's left over. The successor principle we are discussing says:

> For any finitely many X's, there are some Y's that succeed the X's.

To see that the successor principle implies Infinity, suppose there only finitely many things, and let the X's be all of them. If there are any Y's that succeed these X's, then these are included among the X's. It then follows that the X's are in one-to-one correspondence with a proper subplurality of the X's, which would mean that there are infinitely many X's after all.

To see that Infinity implies the successor principle, note that if there are finitely many X's and infinitely many things, then there is some y distinct from each X. Then let the Y's be those things which are either identical to y or else among the X's; it's clear that the Y's succeed the X's.

So—like the mereologist—the number theorist would like to bridge the gap from many to one.

Frege floated a proposal for how to do just this: by providing a *criterion of identity*.

> If we are to use the symbol a to signify an object, we must have a criterion for deciding in all cases whether b is the same as a, even if it is not always in our power to apply this criterion (sec. 62).

In particular, we have such a criterion of identity for numbers—saying what it is for numbers m and n to be the *same* number—in "Hume's Principle":[6]

> **Numbers.** $\#X = \#Y$ iff there are exactly as many X's as Y's: that is, $X \preceq Y$ and $Y \preceq X$.

The proposal is that the principle Numbers is a *definition* of the operator "the number of X's" (abbreviated $\#X$), telling us what individual numbers *are*. As Frege suggestively puts it, "we carve up the content in a way different from the original way, and this yields us a new concept" (sec. 64). In this case, we "carve up the content" of a statement about one-to-one correspondence in a new way, as an identity statement, thus yielding the concept of an individual "self-subsistent" number. We can define a *number* to be whatever is the number of some things: that is,

> x is a **number** iff, for some X's, $x = \#X$.

It's worth noting that Frege ultimately rejected the adequacy of this principle as a definition of number. He instead derived Numbers from another definition which was, by his lights, more basic. This other definition relied on his theory of *extensions*, which is famously vulnerable to Russell's paradox (see Section 3). But more recently *neo-Fregeans* have taken up and defended the proposal that Numbers does not need further justification in terms of some prior definition of number (e.g., Wright 1983). It's this neo-Fregean idea, rather than Frege's original version, which I'll be exploring.

For this to be a suitable definition for mathematical purposes, numbers ought to have the right kind of structure. Numbers inherit their arithmetical structure from the cardinal structure of pluralities. For example, we can define the ordering of numbers thus:

[6] As an anonymous referee helpfully points out, the equivalence between this formulation and the more standard one—which says that $\#X = \#Y$ iff the X's and the Y's are in one-to-one correspondence—depends on the plural form of the Schröder-Bernstein theorem; this again relies on Global Well-Ordering, or some other similarly strong background principle. Once again, these technicalities are bypassed if we restrict the principle to the special case of finite pluralities.

$\#X \preceq \#Y$ iff there are no more X's than Y's: that is, $X \preceq Y$.

(We can prove that this relation is well-defined: if $\#X = \#X'$ and $\#Y = \#Y'$, then $X \preceq Y$ iff $X' \preceq Y'$, by Numbers and the transitivity of the plural $\preceq$ relation.) Using this definition, we can straightforwardly *derive* standard principles about numbers from the corresponding principles of plural logic. For any numbers x, y, and z,

Reflexivity. $x \preceq x$.

Transitivity. If $x \preceq y$ and $y \preceq z$ then $x \preceq z$.

Totality. Either $x \preceq y$ or $y \preceq x$.

Well-Ordering. If there is some number x such that $\varphi(x)$, then there is some number x such that $\varphi(x)$ and for every number y such that $\varphi(y)$, $x \preceq y$. (In short, if there are any φ-numbers, then there is a least φ-number.)

These have exactly the same form as the plural principles, except now they are singular principles about special objects—numbers. Numbers also obey a stronger "individuation" principle than pluralities do:

Anti-Symmetry. If $x \preceq y$ and $y \preceq x$ then $x = y$.

One might have thought that the laws of cardinality were to be derived from the laws governing cardinal numbers. According to this picture (as Frege says about an analogous case about parallel lines and directions) "this is to reverse the true order of things" (sec. 64).

Numbers is an *abstraction* principle. We take a certain respect of *similarity*—standing in one-to-one correspondence—and abstract from it a notion of numerical *sameness* for a special sort of object. The key technical observation of neo-Fregeanism is that the laws of arithmetic follow from Numbers, using suitable definitions (along the lines of the definition of the order for numbers).[7] The key philosophical claim of neo-Fregeanism is that Numbers is true, and moreover an *innocent* truth: it is analytic, or trivial, or a logical truth, or in any case has some special status that makes it a suitable "foundational" truth for the epistemology and metaphysics of mathematics. I'll return to the question of how to make this more precise in Section 3.

[7] As I mentioned earlier, Frege was working in a context of relational higher-order logic (as do most neo-Fregeans), while I've rephrased things in terms of plural logic. One of the annoying discrepancies between these two frameworks is that, while there are standardly empty concepts, like $\lambda x(x \neq x)$, there is not standardly any "empty plurality" of nothing at all. This has the awkward upshot that we don't get the number *zero* from the plural version of Numbers. In order to derive Infinity from Numbers, we also need an extra side premise—that there is at least one non-number.

One nice bonus is that Numbers also answers the other objection to "plural arithmetic" that I raised earlier—that there might not be sufficiently large pluralities to guarantee that every number has a successor. In fact, we can derive Infinity from Numbers, by successively applying Numbers to pluralities of numbers themselves. While it is a nice technical feat, this does raise further worries about whether the principle might be more substantive than it originally sounded. I'll return to this point in Section 3 as well.

From a formal perspective, mereology is exactly analogous to arithmetic. Let's go through the points of analogy.

We have already described mereological structure for pluralities—the plural "covering" relation—which is analogous to the cardinal structure of pluralities which served as our foundation for arithmetic. In the mereological case the structure isn't quite purely logical, since we have used a notion of "ur-parthood" to get things going, though we have not presupposed anything substantive about this relation's structure.

To get from plural mereology to mereology proper, we need an individual ("self-subsistent") thing where we had many things. We can do this with an abstraction principle analogous to Numbers:

> **Sums.** $\sigma X = \sigma Y$ iff the X's and the Y's are mutually covering: that is, $X \leq Y$ and $Y \leq X$.
>
> x is a **mere sum** iff for some X's, $x = \sigma X$.

It will become clearer later why we speak of "mere" sums here. The vague and picturesque idea is that, as numbers are special things whose nature is exhausted by their role in cardinal arithmetic, mere sums are special things whose nature is exhausted by their role in classical mereology.

As in the case of numbers, mere sums inherit mereological structure from the structure of pluralities:

> $\sigma X \leq \sigma Y$ iff the X's are covered by the Ys.

(Again, this is well-defined, because if $\sigma X = \sigma X'$ and $\sigma Y = \sigma Y'$, then by Transitivity, $X \leq Y$ iff $X' \leq Y'$.) It follows from Sums and the two definitions that mere sums satisfy Non-Extensional Mereology. That is, the following principles hold for all mere sums x, y, and z:

> **Reflexivity.** $x \leq x$.
>
> **Transitivity.** If $x \leq y$ and $y \leq z$ then $x \leq z$.

Supplementation. If for any z that overlaps x, z overlaps y, then $x \leq y$.

Fusions. If there is some x such that $\varphi(x)$, then there is some y such that, for any z, z overlaps y iff z overlaps some x such that $\varphi(x)$.

(In this context, "x overlaps y" means that for some mere sum z, $z \leq x$ and $z \leq y$.) These follow straightforwardly from the plural principles by the same names, just as facts about the ordering of numbers followed from the cardinal ordering of pluralities. Mere sums also obey a stronger "individuation" principle than pluralities do:

Anti-Symmetry. If $x \leq y$ and $y \leq x$, then $x = y$.

That is, unlike pluralities, mere sums are *extensional*, in the mereological sense, and thus satisfy all of classical mereology.

In the mereological case we can also go a bit further, by hooking up mere sums with our original ontology: there is a natural sense in which the mere sum of the X's has each of the X's as a *part*. This arises from the natural way of associating each object with a certain mere sum: we can let $\sigma\{x\}$ be the mere sum of the "singleton plurality" of just those things which are x.

$x \leq y$ iff $fx \leq fy$ (with respect to the defined ordering of mere sums), where

$$fx = \begin{cases} x & \text{if } x \text{ is a mere sum} \\ \sigma\{x\} & \text{otherwise} \end{cases}$$

This definition extends the originally defined $\leq$-order for mere sums to also apply to things which are not mere sums. (So even though we have given two definitions of $\leq$ for mere sums, there is no ambiguity, because both definitions agree in that case.) Furthermore, it follows from the definitions that if x and y are not mere sums, and x is an *ur-part* of y, then $x \leq y$. Moreover, this relation is guaranteed to satisfy the axioms of Non-Extensional Mereology.[8] So the end result of our labors is that,

[8] The definition guarantees that x and fx are order-indiscernible: $x \leq fx$ and $fx \leq x$. This makes it easy to carry over ordering properties from mere sums to those things which get mapped to them by f (as long as they don't mention identity). For example, we can prove Fusions as follows. Suppose there is at least one X. Let the fX's be those mere sums which are fx for some x among the X's. By Fusions for mere sums, there is some fusion y of the fX's. Then y is also a fusion of the X's—since anything that overlaps x also overlaps fx, and vice versa.

Here f is a an **equivalence** between the $\leq$-ordering of everything and the smaller $\leq$-ordering of mere sums, in the sense from category theory—an equivalence is essentially an isomorphism, except letting order-indiscernibility playing the role of identity. (Hovda 2016 calls this kind of correspondence a "quasi-isomorphism"; see also Cotnoir 2010.)

given Sums, we can extend our original notion of ur-parthood to a nearly classical parthood relation.

This, then, is the argument for Mereological Universalism—and indeed, for the rest of Non-Extensional Mereology as well. The premise is Sums. Non-Extensional Mereology, including the existence of arbitrary fusions, follows from this premise (using the definition of the extended $\leq$-ordering). The situation is formally exactly analogous to how the arithmetic of cardinal numbers follows by a chain of definitions from the abstraction principle Numbers.

As I mentioned earlier, neo-Fregeanism involves a technical observation—the theory of arithmetic follows from the abstraction principle Numbers—and a philosophical claim—that Numbers is an innocent truth. We have shown that an analogous technical observation also applies to mereology: Non-Extensional Mereology follows from the abstraction principle Sums. So, how about the analogous philosophical claim? Might Sums be an innocent truth, in the same sense?

3 Innocence

Many philosophers have been attracted to the idea that mereology is innocent. David Lewis writes, "given a prior commitment to cats, say, a commitment to cat-fusions is not a *further* commitment" (1991, 81, original emphasis). There are various ways of trying to make good on this idea. Lewis's idea was that the cats are "the same portion of Reality" as their fusion. One daring way of spelling this out is the "Composition as Identity" thesis: the cats are literally identical to their fusion (Baxter 1988; for dissent see Yi 1999; for further discussion see Cotnoir and Baxter 2014). Others are more cautious: Lewis retreats to the view that composition is *analogous* to identity. Agustín Rayo tries to do similar work with *fact*-identities, rather than (singular or plural) *object*-identities: he suggests that "For there to be a table *just is* for there to be some things arranged tablewise" (2013, 3, original emphasis). Others hold that cat-fusions are *grounded* in cats, and that such derivative entities don't count as "further commitments" (deRosset 2010; Cameron 2014). Katherine Hawley (2014) proposes that a commitment to fusions might not "count as extra" in a more modest sense: these further commitments are sufficiently theoretically virtuous, perhaps due to the explanatory work they do, that they don't offend against parsimony.

Neo-Fregeanism offer us a different way of articulating the idea of mereological innocence: perhaps commitment to fusions is innocent in something like the way that *definitions* are innocent commitments.

Consider an axiomatic theory—for concreteness, second-order Peano arithmetic. Each axiom specifies a commitment, a respect in which the theory might turn out to be wrong. It might turn out that some number has no successor, or two numbers might turn out to have the same successor, or the induction axiom might turn out to have counterexamples. In contrast, the theorems which are proved from those axioms are not *further* commitments: any way they might turn out to be false is already a way at least one of the axioms might turn out to be false.[9] Besides axioms and theorems, we also have stipulative *definitions*—such as the definition of "less than" in terms of successor. These are unlike theorems, in that they cannot be straightforwardly proved from the axioms, but they are also unlike further axioms, in that they are not "further commitments" of the theory. They don't introduce extra risks of error, or constraints on applicability. This is despite the fact that some interesting statements, such as "Among any numbers there is a least number", cannot be proved from those axioms alone—that is, they cannot be proved without appeal to definitions.[10]

Neo-Fregean arithmetic is based on the idea that *none* of the truths of arithmetic are "further commitments" in this sense, because in fact they can all be logically derived from definitions alone. The principle Numbers is taken to be the definition of "the number of *X*'s"—though, as Frege puts it, "Admittedly, this seems to be a very odd kind of definition, to which logicians have not paid enough attention" (1953, sec. 63).

Exploring the prospects for the neo-Fregean program would take us too far afield. In any case, I'm far from convinced it's right, myself. But *if* Numbers is innocent, then this also provides us with a way in which Sums—and thus a great deal of mereology—might be innocent.

But I'm afraid I'm not even sure of the conditional: if Numbers is innocent, so

[9] These "might" claims are heuristic—it's difficult to pin down a sense in which they are all true. Clearly metaphysical possibility won't do, if arithmetic consists of metaphysically necessary truths; rather we want some sort of epistemic modality. But also, since we sometimes accidentally produce false proofs, it's plausible that in some epistemic sense the theorems might turn out to be wrong independently of the axioms.

[10] The word "statement" elides some complications. One might think that the *proposition* that among any numbers there is a least number just is the proposition that among any numbers, the *X*'s, there is one that bears the ancestral of successor to each of the *X*'s—that is, a proposition which can be expressed without using "less than", and which can be proved from the standard axioms. (Compare Rayo's "just is" claims.) I think the right thing to say, if propositions are as coarse-grained as this, is that provability is not properly speaking a feature of *propositions*, but rather a feature of ways of presenting propositions—propositions "under a guise"—and the same then goes for features like being an axiom or a definition, or being *innocent*. Talk of "commitments" should be understood in a correspondingly fine-grained way.

is Sums. It doesn't suffice to point out that Sums has the same formal structure as Numbers: for everyone must admit that some principles which have the same formal structure as Numbers and Sums are false. This is the famous "Bad Company" problem (for overview see Linnebo 2007). The original offending instance is Frege's:

> **Basic Law V.** The set of X's is the set of Y's iff each X is a Y, and each Y is an X.

Russell's Paradox shows that this principle is not only false, but inconsistent, given standard plural logic.[11]

Fortunately, both Numbers and Sums are consistent—so they don't fall prey to the particular problem that Basic Law V faces (Boolos 1987). But consistency is also not enough for innocence: some abstraction principles are logically consistent, but still false. We can see this because different individually consistent principles are logically inconsistent with each other. For example:

> **Nuisance.** $nX = nY$ iff only finitely many X's are not Y's, and only finitely many Y's are not X's.

This principle is consistent, but it implies that there are only finitely many things. Since Numbers implies Infinity, Numbers and Nuisance are not both true (Boolos 1990).[12]

The form of an abstraction principle can't guarantee its innocence. If any abstractions are innocent, we need some further criterion to distinguish them. Unfortunately, there is no agreed-upon criterion.

Wright (1997) proposed the test of *conservativeness* (in the sense of Field 1980; see discussion in Wright 1999, sec. 2.5; Weir 2003, 22–23). The idea is that a legitimate abstraction principle should not only be consistent, but also consistent with arbitrary "empirical" hypotheses. Here is a way of articulating this idea. Suppose A is an

[11] In our context (in which we have no "empty plurality") this also relies on the auxiliary assumption that there is at least one non-set. Let x be a *self-member* iff for some X's, x is the set of the X's and x is one of the X's. If there is at least one non-set, then there is at least one non-self-member. In that case the set of non-self-members is a self-member iff it is not a self-member.

[12] By "consistent" I mean *semantically* consistent, throughout—that is, a consistent sentence or theory is one that has a (full) model. It's worth noting that I have implicitly relied here on the principle that no inconsistent theory is true. This general principle is more controversial than it might sound, given this standard technical gloss on consistency. Indeed, applied to semantic consistency for second-order logic, it amounts to a very powerful reflection principle, called Kreisel's Principle (Shapiro 1987). (Briefly, the general principle requires that there are set-theoretic models which are "big enough" to model the truths about *all sets*.) Even so, I don't think the instances of this principle these arguments rely on should be especially controversial, since the inconsistencies involved are pretty straightforward.

abstraction principle, of the form "$\alpha X = \alpha Y$ iff $\varphi(X, Y)$". (Here α is an operator taking plural terms to singular terms: we substitute # in the case of Numbers, or σ in the case of Sums.) If T is a theory, let T^α be the result of restricting the quantifiers in T to non-abstracts. (That is, we systematically replace each occurrence of $\forall x(\dots)$ with $\forall x(\forall Y(x \neq \alpha Y) \to \dots)$.) Then:

> A is **Field-conservative** iff for any theory T in which α does not occur, if T is consistent, then T^α and A are jointly consistent.

The Nuisance principle is not Field-conservative, because it restricts the size of the universe: it implies, for instance, that there are not infinitely many space-time points. In contrast, Numbers is Field-conservative. Numbers implies that there are infinitely many things—but this is only because it implies that there are infinitely many *numbers*. It is consistent with the hypothesis that there are only finitely many non-numbers.

Like Numbers, the Sums principle is Field-conservative.[13] So this is a point in favor of its being one of the good abstraction principles, rather than the bad.[14]

Unfortunately, conservativeness is also not a sufficient condition for the truth of an abstraction principle: for again, there are pairs of individually *conservative* abstraction principles which are jointly inconsistent, though examples are a bit more technical than the principles we've considered (see Weir 2003, 27). We could pursue this line further, looking for better candidates for sufficient conditions for an abstraction principle to be innocent, but the going gets hard, and as I understand it, no especially compelling candidate has been struck on so far (for overview see Linnebo 2007, sec. 3; Heck 2011, ch. 10 postscript). As far as we've investigated, Sums is holding even with Numbers, in contrast with vicious abstraction principles like

[13]To show this, it suffices to show that any model of T can be embedded as a substructure of some model of Sums, where the domain of the original model is mapped onto the non-abstracts. Suppose we have a model of T with domain D and an ur-parthood relation P. Let $(\Sigma, \leq)$ be the set of equivalence classes of subsets of D modulo mutual covering, with the covering relation defined in terms of P. This is a model of classical mereology. Then we can consider a model whose domain is $D \cup \Sigma$, which extends ur-parthood to $P \cup \leq$, and for any sets $X \subseteq D$ and $Y \subseteq \Sigma$, let $\sigma(X \cup Y)$ be the unique fusion in Σ of the equivalence class of X together with Y. It is straightforward to check that this is a model of Sums.

[14]It's worth noting an alternative moral one might draw from this technical fact. Field (1980) used the conservativeness of arithmetic as a way of resisting the claim that mathematics is *indispensable* for characterizing the physical world—and thus as a way of defending the view that there are no numbers at all. The conservativeness of mathematics provides a way of explaining why *supposing* it to be true is indulging in a harmless fiction. Analogously, one might deploy the conservativeness of Sums as part of a defense of the view that there are no mereological sums, and yet the fiction that there are is a harmless one (cf. Rosen and Dorr 2002).

Basic Law V or Nuisance. But there is clearly still further to go.[15]

But we also ought to touch on two important ways in which the Sums principle is not like Numbers. One of these counts against Sums—it is a respect in which Sums may be *less* innocent than Numbers. The other counts in Sums' favor—it is a respect in which Sums may be *more* innocent than Numbers.

The right-hand-side of the Numbers principle is statable using just logical vocabulary.[16] In contrast, the right-hand side of Sums uses the notion of "covering", which was defined in terms of "ur-part", which is not a logical constant. Some hold that what innocence Numbers may possess is owed to its logicality (e.g. Wright 1999, 16; Fine 2002; Heck 2011, 235–6). If this is indeed a necessary condition for an abstraction principle to be innocent, then this is a way the conditional—if Numbers is innocent, so is Sums—may fail.

Restricting attention to logical abstraction principles spares the neo-Fregean some technical pain, but it seems to me that this restriction comes at a high cost. By my lights, an especially appealing feature of neo-Fregeanism is its purported power to *unify* the foundations of arithmetic together with other cases of abstraction, both mathematical and otherwise, including colors, sentences, truth-values, lengths, shapes, and directions. (Frege used the last two examples as motivation.) Throughout mathematics, passing to a *quotient* is a pervasive technique, with just this structure: a certain respect of similarity—in general, a *congruence* of certain structure—gives rise to a new sort of mathematical object for which the congruence relation provides a criterion of identity. One of the delightful possibilities Frege raises is that the relationship between numbers and cardinality is of the same kind as the relationship between directions and parallelism, or between a circle and rays in a punctured plane, or between the cyclic group $\mathbb{Z}/n$ and congruence modulo n, or between the fundamental group of a topological space and smoothly deformed curves.[17] These other applications of abstraction are non-logical. It would be a shame if technicalities require giving up this insight of commonality. I don't think neo-Fregeans should give up on non-logical abstraction just yet.

[15] The task faces the in-principle obstacle that the question of which abstraction principles are true is undecidable: any general procedure for deciding which abstraction principles are true would provide a procedure for deciding whether arbitrary statements of arithmetic are true (see Heck 2011, ch. 10).

[16] Recall that the original version used relational higher-order logic, instead of plural quantification over pairs; but for simplicity we're counting ordered pairs as honorarily logical.

[17] It's worth noting that orthodox mathematics based on set-theoretic foundations breaks this connection. The now-standard way of constructing quotient structures uses the method of equivalence classes (just as Frege proposed). But this approach won't work when it comes to numbers: in standard set theory, in general there are no equivalence classes under equinumerosity (except in the case of the class containing just the empty set). For instance, there is no set of all one-membered sets.

There is also a respect in which the theory of mereology based on Sums has a considerable advantage over a theory of arithmetic based on Numbers. One reason some philosophers have been suspicious of principles like Numbers is that they are *impredicative*, in the sense that they quantify unrestrictedly over individuals which include numbers themselves (see Linnebo 2007, sec. 3.4). But how then could such a principle be a way of introducing the concept of a number? It would be appealing to instead use a restricted principle, which only uses quantifiers that are restricted to non-numbers (or more generally, to non-abstracts).[18] Any so-restricted abstraction principle is consistent. But this sort of principle is substantially weaker than the unrestricted version. In particular, a version of Numbers which does not guarantee the existence of numbers of numbers won't suffice to derive Infinity—and thus won't be strong enough to imply all of arithmetic.

In contrast, though, a "predicatively" restricted version of Sums, while on its face weaker than the full-fledged Sums principle, is still strong enough for mereology. In order to derive Non-Extensional Mereology from Sums, we have no need for "mere sums of mere sums". The basic reason for this is that a fusion of fusions of X's is also a fusion of X's—whereas a number of numbers of X's need not be a number of X's (and likewise a set of sets of X's need not be a set of X's). So in this respect, "neo-Fregean mereology" is in *better* standing than neo-Fregean arithmetic, because a predicatively restricted abstraction principle is perfectly adequate for this purpose.

Sums shares this feature with the other applications of abstraction I just mentioned: truth-values, directions, fundamental groups, and so on, don't rely on impredicative abstraction. Numbers is the unusual case. And even in the case of arithmetic, if we already know on some other grounds that there are infinitely many things, then predicatively restricted Numbers is good enough for arithmetic after all. So while it is less technically exciting than the full-fledged version, restricting abstraction to non-abstracts holds promise.

Finally, let's examine some principles whose innocence would *not* be vindicated, even if abstraction is innocent. As we noted earlier, Sums has the consequence that *mere sums* obey classical mereology, but the mereology that applies to arbitrary things need not be extensional. So, for example, the resulting mereology need not satisfy this principle

Uniqueness. If x fuses the X's and y fuses the X's, then $x = y$.

(Recall that x **fuses** the X's iff, for any z, z overlaps x iff z overlaps some X.)

[18] This should be contrasted with a different proposal (from Dummett 1991) which involves adding a predicativity restriction to the *background plural logic*.

Other approaches to the "innocence" of mereology treat the existence of fusions and the uniqueness of fusions on a par (e.g. Varzi 2014). For instance, the slogan that composition is closely analogous to identity strongly suggests a Uniqueness-friendly picture (applying Euclid's principle that "things which equal the same thing also equal one another"). In contrast, if the innocence of mereology is the kind of innocence that abstraction principles are supposed to enjoy, then it does not encourage this picture at all. The basic reason is that abstraction gives us a cheap way to "expand" our ontology with new things standing in appropriate relations to the old—which is all that the existence of fusions requires. But it doesn't give us any way to "collapse" our ontology—which the *uniqueness* of fusions might require.

As an example, suppose that a certain statue, Alfred, and a certain piece of clay, Clay, are each parts of one another (Thomson 1998, 155). Uniqueness is in conflict with this picture. No matter what else there may be in the universe, and no matter how parthood is distributed (so long as it is transitive) whatever shares a part with Alfred will also share that same part with Clay, and vice versa. So Clay fuses the singleton plurality {Clay}, and Alfred also fuses {Clay}, and yet Clay and Alfred are distinct. To get Uniqueness from this statue-clay picture, we would have to somehow collapse Alfred and Clay into a single object. We can get the existence of arbitrary fusions simply by adding things in—or as Frege put it, by "carving up our content" in such a way as to reveal new kinds of object. But the same process won't eliminate objects we have already exposed, like Alfred and Clay.

Part of how we justified the claim that Sums might be innocent as definitions are was by pointing out that Sums is *conservative* (in Field's sense). This means, in particular, that Sums is consistent with any distribution of ur-parthood over things that are not mere sums. But Uniqueness is not conservative in this sense: Thomson's theory of part, for instance, cannot be recovered as a restriction of an extensional ontology.

Formally, there is a short route from Sums to extensionality: we simply need to posit that *everything* is a mere sum—in particular, we might hold that everything is the mere sum of itself:

$$\sigma\{x\} = x$$

In fact, the combination of this claim with Sums is equivalent to full-fledged classical extensional mereology (for the whole universe, not just mere sums). But even if this claim is true, it is hard to see how it could be *innocent* in the same way that Sums might be. Compare the analogous principle that $\#\{x\} = x$. This, in combination with Numbers, has the drastic consequence that there is at most one thing (since $\{x\}$ and $\{y\}$ are in one-to-one correspondence, for any x and y).

Abstraction presents us with a distinctive account of how mereology might be innocent, which is quite different from the thesis that composition is analogous to identity.[19] I hope further exploration of this link between the metaphysics of composition and the philosophy of mathematics will prove fruitful.

References

Baxter, Donald L. M. 1988. "Identity in the Loose and Popular Sense." *Mind* 97 (388): 575–82. http://www.jstor.org/stable/2255191.

Boolos, George. 1987. "The Consistency of Frege's Foundations of Arithmetic." In *On Being and Saying: Essays in Honor of Richard Cartwright*, edited by J. Thomson, 3–20. MIT Press.

———. 1990. "The Standard of Equality of Numbers." In *Meaning and Method: Essays in Honor of Hilary Putnam*, edited by George Boolos, 261–77. Cambridge University Press.

Cameron, Ross P. 2014. "Parts Generate the Whole, but They Are Not Identical to It." In *Composition as Identity*, edited by Aaron J. Cotnoir and Donald L. M. Baxter. Oxford University Press.

Cotnoir, Aaron J. 2010. "Anti-Symmetry and Non-Extensional Mereology." *Philosophical Quarterly* 60 (239): 396–405.

Cotnoir, Aaron J., and Donald L. M. Baxter, eds. 2014. *Composition as Identity*. Oxford: Oxford University Press.

deRosset, Louis. 2010. "Getting Priority Straight." *Philosophical Studies* 149 (1): 73–97. doi:10.2307/40606343.

Donaldson, Thomas. 2016. "The (Metaphysical) Foundations of Arithmetic?" *Noûs*, May, 1–27. doi:10.1111/nous.12147.

Dummett, Michael A. E. 1991. *Frege: Philosophy of Mathematics*. Cambridge, Mass:

[19] I think the abstraction approach may be compatible with some of the other "innocence" proposals that have been floated. For instance, the grounding theorist might want to say that abstracts are derivative: perhaps the number of planets is *derivative* from the planets themselves. But Donaldson (2016) raises serious problems for the ground-theoretic version of Numbers: it gives rise to cycles and non-well-founded chains of grounds. I suspect that similar issues also arise for the ground-theoretic version of Sums (in its full "impredicative" version), though I have not worked this out.

It is, by my lights, more plausible that definitions are an instance of the kind of fact identities that Rayo discusses. If Numbers is a definition, then plausibly for $\#X$ to be $\#Y$ *just is* for the X's and Y's to be in one-to-one correspondence (see Rayo 2013, 76ff.). The analogous point would also apply to Sums. Perhaps we should endorse the identity form of the abstraction principle: for σX to be σY *just is* for the X's and Y's to be mutually covering. This would be a way of spelling out Rayo's suggestion that "For there to be a table *just is* for there to be some things arranged tablewise."

Harvard University Press.

Field, Hartry H. 1980. *Science Without Numbers: A Defence of Nominalism*. Princeton University Press.

Fine, Kit. 2002. *The Limits of Abstraction*. Oxford University Press.

Frege, Gottlob. 1953. *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*. Translated by J. L. Austin. 2d rev. Oxford: Blackwell.

Hawley, Katherine. 2014. "Ontological Innocence." In *Composition as Identity*, edited by Donald L M Baxter and Aaron J. Cotnoir. Oxford: Oxford University Press.

Heck, Richard G. 2011. *Frege's Theorem*. 1 edition. Oxford: Oxford University Press.

Hovda, Paul. 2014. "Natural Mereology and Classical Mereology." In *Mereology and the Sciences*, edited by Claudio Calosi and Pierluigi Graziani, 141–59. Cham: Springer International Publishing. http://link.springer.com/10.1007/978-3-319-05356-1.

———. 2016. "Parthood, Plurality, and Sets." University of Southern California.

Inwagen, Peter van. 1994. "Composition as Identity." *Philosophical Perspectives* 8: 207–20. http://www.jstor.org/stable/2214171.

Lewis, David. 1991. *Parts of Classes*. Cambridge: Blackwell.

Linnebo, Øystein. 2007. "Introduction." *Synthese* 170 (3): 321–29. doi:10.1007/s11229-007-9267-5.

———. 2010. "Pluralities and Sets." *Journal of Philosophy* 107 (3): 144–64.

———. 2014. "Plural Quantification." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2014. http://plato.stanford.edu/archives/fall2014/entries/plural-quant/.

Rayo, Agustin. 2013. *The Construction of Logical Space*. Oxford: Oxford University Press.

Rosen, Gideon, and Cian Dorr. 2002. "Composition as Fiction." *The Blackwell Guide to Metaphysics*, 151–74. http://users.ox.ac.uk/~sfop0257/papers/CompositionAsAFiction.pdf.

Shapiro, Stewart. 1987. "Principles of Reflection and Second-Order Logic." *Journal of Philosophical Logic* 16 (3): 309–33. doi:10.1007/BF00152934.

Thomson, Judith Jarvis. 1998. "The Statue and the Clay." *Noûs* 32 (2): 149–73.

http://www.jstor.org/stable/2671962.

Varzi, Achille C. 2014. "Counting and Countenancing." In *Composition as Identity*, edited by Aaron J. Cotnoir and Donald L. M. Baxter, 47–69. Oxford University Press.

Weir, Alan. 2003. "Neo-Fregeanism: An Embarrassment of Riches." *Notre Dame Journal of Formal Logic* 44 (1): 13–48. doi:10.1305/ndjfl/1082637613.

Wright, Crispin. 1983. *Frege's Conception of Numbers as Objects*. Aberdeen University Press.

———. 1997. "On the Philosophical Significance of Frege's Theorem." In *Language, Thought, and Logic: Essays in Honour of Michael Dummett*, edited by R. Heck, 201–44. Oxford University Press.

———. 1999. "Is Hume's Principle Analytic?" *Notre Dame Journal of Formal Logic* 40 (1): 6–30. doi:10.1305/ndjfl/1039096303.

Yi, Byeong-Uk. 1999. "Is Mereology Ontologically Innocent?" *Philosophical Studies* 93 (2): 141–60.