COUNTING IN GOODMAN AND QUINE'S CONSTRUCTIVE NOMINALISM

ALEXANDER R. PRUSS AND EMIL JEŘÁBEK

Abstract. It is well-known that one cannot use first-order logic with identity and the predicates $\mathrm{Cat}(x)$ and $\mathrm{Dog}(x)$ to say that there are more cats than dogs. Nonetheless, Goodman and Quine (1947) offered an ingenious translation of the sentence into a richer but thoroughly finitist and nominalist language with mereological vocabulary and size comparison for individuals. However, their translation as it stands fails in the case of counting comparisons involving overlapping objects (say, conjoined twin cats). Furthermore, we prove that no general translation of equinumerosity (and hence of "more") can be given in the overlapping object setting using the predicates in Goodman and Quine's translation, assuming size comparison can be cashed out by counting mereological atoms, and we use computational complexity theory to prove a more general inexpressibility result. We end with some open questions.

1. Introduction

Goodman and Quine in "Steps towards a constructive nominalism" [3] have made one of the most valiant attempts ever at giving a reduction of logical meta-theory and ordinary counting language to an uncompromising physicalist "constructive" nominalism with no cheating. There are no properties, tropes, sets, classes, or types, and there is no cheating with second-order quantifiers or infinite sentences. All we have are physical individuals

Date: This is a post-peer-review, pre-copyedit version of an article to be published in *Synthese*. The final authenticated version is not yet available online.

governed by classical mereology, with the theory explicitly developed to be compatible with finitism—the hypothesis that there are only finitely many individuals.

We will focus on Goodman and Quine's clever account of counting comparison sentences such as "There are more cats than dogs", assuming finitism. We show that Goodman and Quine's own account fails given overlapping individuals such as conjoined twins, and then go on to prove that given their mereological commitments to overlapping individuals, there is no way to extend their account to handle such objects without going beyond the theoretical resources they used for their cats-and-dogs example.

Next, we observe that there are some additional nominalistically acceptable resources that Goodman and Quine did not draw on for their cats-and-dogs account, but prove that within the scope of finitism these resources are still insufficient if we understand the translation task as one of finding a schematic formula that says "The number of objects x such that $F(x)$ equals the number of objects x such that $G(x)$" that works for any predicates F and G. We then discuss some open problems related to possible nominalist solutions for Goodman and Quine coming from either relaxing the finitism of "Steps" or relaxing the details of the translation requirement.

2. Counting and "bigger"

2.1. **Cats and dogs.** In First-Order Logic with identity and predicates $\mathrm{Cat}(x)$ and $\mathrm{Dog}(x)$, for any specific finite n we can say that there are at most, at least or exactly n cats and/or dogs by an appropriate quantified statement. For instance, that there exactly two cats can be said as

$$\exists x \exists y (\mathrm{Cat}(x) \wedge \mathrm{Cat}(y) \wedge x \neq y \wedge \forall z (\mathrm{Cat}(z) \to (z = x \vee z = y))).$$

Van Benthem and Icard [11] call this "counting in the syntax". But it is well known that we cannot in this way say things like "There are finitely many cats"[1] or "There are more cats than dogs."[2] With infinite sentences, of course, we could say that there are more cats than dogs by saying that there is at least one cat and no dogs, or at least two cats and at most one dog, or at least three cats and at most two dogs, ... But that won't fit with the finitism of "Steps".

Goodman and Quine, however, have an additional resource available, namely mereology, and give an ingenious account of how to say that there are more cats than dogs. Stipulate that a *bit* is a part of a cat or a dog that is the same size as the smallest of the cats and dogs. Then stipulate that:

BITS There are more cats than dogs if and only if every individual that has a bit of every cat is bigger than some individual that has a bit of every dog.

(What "size" and "bigger" could be taken to mean will be discussed in Section 2.2.)

But now suppose that there are two cats and one dog, while (a) the two cats share a leg, (b) the dog is smaller than that shared leg (say, it's a fetal dog), and there is no other overlap. Then consider any individual C that is a bit of the shared cat leg, and any individual D that has a bit of the dog. Then C is an individual that has a bit of every cat, and yet D is at least

[1]The impossibility of saying this follows immediately from compactness, since any sentence ϕ translating "There are finitely many cats" will be compatible with any finite number of sentences ϕ_n saying that there are at least n cats, and hence there will be a model where ϕ and all the ϕ_n are true.

[2]For a simple elementary proof, see [**?**].

as large as C, so Bits delivers the incorrect verdict that there are no more cats than dogs.

Now, while conjoined twins are rare, nonetheless it is a central part of Goodman and Quine's theory that objects can overlap. For, famously (or infamously), in order to avoid having to posit linguistic types, they suppose that the world is all filled with invisible linguistic tokens. For instance, if you are reading this paper on a screen, on the white margins of the page, there are pixels that spell out your grandparents' names. You don't see these pixels because they are surrounded by other pixels of the same color. If you turned these surrounding pixels a different color, the pixels spelling out your grandparents' names would stand out, but they are there anyway. However, in the same part of the margin where there are pixels spelling out your grandparents' names, there are pixels spelling out Schrödinger's equation, and some of the pixels of your grandparents' names are reused for Schrödinger's equation. And the same is true, but with chunks of white paper instead of pixels, if you are reading this article in hard copy. Thus, Goodman and Quine's world is full of *overlapping* linguistic tokens.

We can fix up Bits to work with conjoined twins. For each cat or dog, say that a part is *unshared* provided that it does not overlap any other cat or dog. Let the unshared portion of a cat or dog be the fusion of all of its unshared parts. Then say that a *bit** of a dog or cat is a part that is the same size as the smallest unshared portion of a cat or dog. Then we can say:

Bits* There are more cats than dogs if and only if every individual that has an unshared bit* of every cat is bigger than some individual that has a bit* of every dog.

But while Bits* works well for conjoined twin cats or dogs in our world, it doesn't work for more complex cases of overlap. Imagine that we have a cat c that is a proper part of a larger cat C. This is problematic for this literal case[3] but for linguistic tokens the phenomenon of one token being a proper part of another is common. Then c has no bits*, since c has no unshared parts, and thus there is no individual that has an unshared bit* of every cat. Therefore, the right-hand-side of Bits* is vacuously satisfied no matter how many dogs and cats there are—in particular we get a counterexample if c and C are the only cats and there are two or more dogs.

One might hope that some other clever definition of bits will do the job, say a definition cleverly making use of spatial continuity[4], so that there is an individual x that is a fusion of pairwise non-overlapping bits†, with each of them a part of a different cat and every cat having exactly one bit† contributing to x, and we can then compare such individuals x in size to individuals made of bits† of dogs. In the case of organisms like cats and dogs we have a hope of doing this, even if there is total overlap like in the c and C example. But we should not expect to be able to do this in full generality—and Goodman and Quine's ontology includes the full generality of a classical mereology. For suppose there exist n distinct mereological atoms $x_1, ..., x_n$, and define a "shcat" as any part of the fusion X of $x_1, ..., x_n$. Then there are $2^n - 1$ shcats, and if $n > 1$ then there is no way to make an individual that contains a bit† of every shcat with no pairwise overlap between the

[3]Though it is worth noting that Kingma [8] has claimed that a fetus is a part of its mother.

[4]We are grateful to an anonymous reader for this suggestion.

bits†, since the fusion of the shcats cannot be partitioned into more than n pairwise non-overlapping parts, while $2^n - 1 > n$ for $n > 1$.

One might try to creatively tweak Bits* to get around worries with more radical overlap. But as we are about to see, if our only resources are mereological predicates and something like "Bigger", this can't be done.

2.2. **Atomic cosmoses.** Say that a mereological cosmos[5] is *atomic* provided that every individual (and hence every part of every individual) has an atomic part—a part with no proper parts. Atomic cosmoses are epistemically possible—for all we know, we inhabit one. And while some philosophers defend the possibility of "gunky" cosmoses with no atoms (for a survey, see [5]), it is plausible that that atomic ones are at least metaphysically possible. Furthermore, any finite mereological cosmos is atomic, since in a non-atomic cosmos there is a downward regress of proper parts, and Goodman and Quine's project in "Steps" is a finitist one. Thus for the sake of the success of their project, the Goodman and Quine account had better work for finite atomic cosmoses.

Note that assuming classical mereology, in an atomic cosmos, if x and y are distinct individuals, then they differ in an atomic part.[6] It follows that we can identify individuals with the fusions of their atomic parts.

[5]The term "cosmos" will be restricted to mereological realities, while the term "universe" will be used in the model-theoretic sense for the set of objects in an abstract model.

[6]If x and y are distinct, then at least one is not a part of the other, or else we violate the reflexivity of parthood. Without loss of generality, x is not a part of y. By Strong Supplementation, there is a part of z of x that doesn't overlap y, and so x and y will differ at least by z.

Atomicity can help us solve two technical problems that afflict Bits in the case of non-overlapping objects. First, if there are infinitely many cats or dogs (or at least some other kinds of things), it could be that for any part x of a cat or dog there is another cat or dog smaller than x, and there is no such thing as a "bit" of a cat or dog such that every cat or dog is at least of its size.

Second, it is unclear how to understand "bigger" in the case of scattered objects, i.e., objects that occupy a disconnected spatial region. Suppose there are 1000 cats and 1001 dogs, but the dogs are all in one building while the cats are scattered all over the earth. Let CatBit be an object consisting of a bit of every cat and DogBit be an object consisting of a bit of every dog. Then there is an intuitive sense in which CatBit is a cloud of cat bits about the size of the earth, while DogBit is a denser cloud of dog bits about the size of a building, so CatBit seems bigger than DogBit.

Maybe the thing to say is that size is volume, and the total volume of DogBit is bigger than the total volume of CatBit, where only the volume within the bits counts, not the volume between them. But now we can't count objects with zero volume, such as point particles would be, since any finite collection of such objects still has zero volume, and point particles are clearly logically possible. Furthermore, the main alternative to point particles are quantum particles whose "position" can be identified with the region of points in space where a measurement might find a particle. But such quantum particles can overlap very significantly. While fermions (say, electrons) cannot share the same position and all the same characteristic such as rest mass, charge, spin and fermionic type because of the anticommutation of their wavefunction, fermions differing in their characteristics (say, two electrons with different spins) can have the same position distribution, and there are no restrictions on sharing position distribution for bosons (say, photons), whether with the same or with different characteristics.[7] Volume won't help with counting objects that are fusions of co-located particles. Or we might try comparing mass instead of size, but then we won't be able to count massless particles, like photons.

In an atomic cosmos, however, the best version of "bigger" seems to be that an object is bigger just in case it has more (mereological) atoms. The BITS account simplifies, as we can replace bits by atoms, and so the problem of a lack of a non-zero lower bound on the size of a cat or dog disappears, and have an unambiguous account of the size of scattered objects.

One might think that comparing sizes by counting atoms is cheating, since the point of the project under examination is to give an account of counting. However, any compunction we might have about comparing the size of objects by counting atoms would also affect comparing the size of object by volume—if there is no philosophical problem with adding up the volumes of bits, there should be no problem with adding up the counts of atoms. The task Goodman and Quine have set themselves is to count complex individuals, and that task is a serious *prima facie* challenge, because the very formulation of the problem seems to presuppose abstract entities, namely sets of objects to be counted. On the other hand, if we are counting the atoms in a complex object, there is no such difficulty: instead of talking

[7]The philosophical literature often discusses co-location in connection with bosons with the same characteristics. As Paul [**?**] notes, in the alleged case of indiscernible co-located particles one can resist the claim that one genuinely has distinct objects. But this resistance is less plausible in the case of discernible particles.

of abstract sets of objects, we just talk of the concrete wholes made of the atoms, and suppose relations such as *more*, *less* or *equal count* between them.

Furthermore, if we allow a physical possibility operator and spatiotemporal arrangement, we have a good chance of being able to *define* the atomic sense of "bigger". Suppose that the actual world's mereological atoms are small enough to each fit within a one millimeter cube.[8] Then:

MOVE Object x is (atomically) bigger than object y if and only if it is physically possible to move the atoms of x and y in such a way that none are destroyed and:

(i) no two atoms of x are within a meter of each other,
(ii) no atom of x has two atoms of y within a centimeter of it,
(iii) every atom of y has an atom of x within a centimeter of it,
(iv) there is an atom of x that has no atoms of y within a centimeter of it, and
(v) each individual atom continues to fit within a one millimeter cube.

Given mild geometric assumptions (say, that we are working in a metric space), this will clearly be correct if x and y do not overlap. But the case of overlap is also unproblematic, as long as we remember that any atom of y that is also an atom of x always has an atom of x within a centimeter of itself—namely it has itself within a centimeter of itself!

Or, alternately, we can proceed counterfactually:

[8]One might think that it's impossible for a mereological atom to occupy more than a point of space, in which this assumption is trivially true. But if extended simples are possible, this is a substantive assumption.

TRASH Object x is bigger than object y where neither object is actually in a trash bin if and only if repeating the procedure of putting an atom of x in the trash at the same time as one puts an atom of y in the trash would eventually result in all the atoms of y being in the trash while not all the atoms of x are yet in the trash.

Again, the case of overlap is unproblematic—it just means that sometimes putting an atom of x in the trash will constitute putting an atom of y there.

Both MOVE and TRASH in effect describe a bijection between collections of atoms.[9] One might wonder at this point whether some similar physical procedure could be used to define counting for non-atomic objects. In the case of non-overlapping objects, this is likely to be the case[10], but in the case of overlapping objects, like our hypothetical cats where one cat is a proper part of another, this is unlikely. For instance, overlapping objects in general are not separable in a way analogous to (i) in MOVE, while putting one cat in the trash will also put in the trash all the cats that are proper parts of it, and putting two objects into trash also puts their fusion into the trash.

As an alternative to counting atoms, one might try counting points occupied by the object, so that x is bigger than y if and only if the set of points occupied by x has bigger cardinality than the set of points occupied by y.[11] But in an atomic cosmos where each atom occupies a single point and no two atoms are co-located, this yields a version of "bigger" equivalent to the atom-count one. Now, one of the main results of our paper

---

[9]We are grateful to an anonymous reader for this observation.

[10]Though we may need to distort the objects—perhaps discontinuously—if, say, one object surrounds another.

[11]We are grateful to a referee for the suggestion.

will be that an atom-count "bigger" cannot be used to define "more" for all finite atomic cosmoses. But any finite atomic cosmos is mereologically isomorphic[12] to a finite atomic cosmos where each atom occupies a single point and no two atoms are co-located, since on classical mereology any two atomic cosmoses with the same number of atoms are mereologically isomorphic. Thus if point-count "bigger" can be used to define "more" for all finite atomic cosmoses, atom-count "bigger" can be used to define "more" for all finite atomic cosmoses without atomic co-location, and hence by isomorphism also for all finite atomic cosmoses. Thus in this context there is no advantage to counting points over counting atoms. And there is a disadvantage: as discussed earlier in connection with quantum mechanics, it appears metaphysically possible for multiple atoms to occupy the same point, and in atomic cosmoses where all the atoms occupy one and the same point, the point-count version of "bigger" never obtains between two objects, and hence is of no help in counting complex objects (which on classical mereology will coincide with the fusions of the co-located atoms).[13] Thus in the

[12]A mereological isomorphism is a bijection between sets of individuals that preserves parthood.

[13]A referee notes an interesting objection. Time-travel suggests the possibility of multiply-located particles [**?**]. But there are practical contexts where we may want to count multilocated instances of the "same" particle as separate individuals due to separate causal efficacy, and then point-counting appears to give a better answer to "how many" questions than atom-counting. However, this conflicts with intuitions that co-located particles (at least of different types) should be counted separately. If one is convinced that multilocated instances should count as separate individuals, a compromise option is to have a mereological ontology where the atoms are particle-types-at-locations, perhaps with "fundamental intensive properties" as in [**?**, p. 57] to handle cases of co-located same-type particles. One will still have co-location of particles of different types, and so defining

subsequent sections of the paper, we will focus on the atom-count version of "bigger".[14]

Finally, it is worth noting that if extended simples are possible, then there is an argument that if atom-count "bigger" cannot be used to define "more", total volume "bigger" cannot be used either, even if we restrict our attention to cosmoses all of whose atoms have non-zero volume. To see this, consider finite atomic cosmoses where each atom is an extended simple with the same non-zero volume and no two atoms overlap spatially. For such cosmoses, total-volume "bigger" is equivalent to atom-count "bigger", and so if total-volume "bigger" can be used to define "more", so can atom-count "bigger" in such cosmoses. And by mereological isomorphism it could then be used to define "more" in all finite atomic cosmoses, which we will argue is impossible.

2.3. **An initial impossibility result.** Atomic mereological cosmoses can be nicely modeled within a monadic second-order (MSO) logic. In MSO, we

"bigger" in terms of point-count will be uninformative in universes where there is just one point occupied by multiple atoms of different types, and atom-count is more promising. In any case, atom-counting is more metaphysically neutral than point-counting, since even on the controversial view where co-location is impossible and multilocated instances should count separately, one can simply use atom-counting with a mereology on which the atoms are occupied points.

[14]One may worry that the atom-count version of "bigger" has some counterintuitive consequences in cases of co-location of atoms. For instance, an object consisting of a million co-located atoms is atom-count "bigger" than an object consisting of a thousand atoms that are spread out through a one-meter ring. This is true, but there is no need to insist that "bigger" should match ordinary language, especially given the argument that if "more" can be defined using a point-count "bigger", then it can be defined using an atom-count "bigger".

have quantification over first-order elements, whose variables are denoted by lower-case letters like x, and over second-order monadic or unary entities, denoted by upper-case letters like X that allow us to form formulas like $X(x)$. We can think of these second-order entities as sets, or properties, or predicates. We also have ordinary predicates with argument slots typed to specify whether they take a first-order or second-order argument. A model of MSO is a tuple $M = (U, \sigma, I)$ where U is the universe—the set of first-order entities—while σ is the signature or collection of predicates and names (for simplicity, we won't have functions), and I is an interpretation of the predicates and names.

Given an atomic mereological cosmos, we can imagine mathematical objects, which we will call *points* (not necessarily spatial ones), that are in one-to-one correspondence with the atoms. If α is an atom, we let α^* be the point corresponding to α. An atomic mereological cosmos's individual β can then be modeled as the non-empty set $\beta^\dagger = \{\alpha^* : \alpha \sqsubseteq \beta\}$ of points, where $\sqsubseteq$ is parthood. Thus, a mereological atom α can be modeled in two different ways in the MSO model M: as a point (or element) α^* of the model's universe U or as a singleton subset $\alpha^\dagger$ of U. Since mereological atoms are a kind of mereological individual, and complex individuals are modeled as subsets of U, for the sake of uniformity it is natural to model an atom α as the singleton $\alpha^\dagger$.

We assume that the mereological axioms are precisely such as to ensure that the above correspondence between mereological individuals in an atomic cosmos and non-empty sets is a bijection with $\alpha \sqsubseteq \beta$ if and only if $\alpha^\dagger \subseteq \beta^\dagger$.

There is one set in the model M that does not correspond to an object in the mereological cosmos: the empty set $\varnothing$. We can treat the empty set as a notational convenience, as it does not add any expressive power on the MSO side. For instance, we can split a universal quantification $\forall X\,\psi$ (existential quantifications can be defined in terms of universal ones by De Morgan) over all sets X in MSO into a conjunction between universal quantification over non-empty sets and evaluation of ψ at the empty set:

$$\forall X(\exists x\, X(x) \to \psi) \wedge \psi'$$

where ψ' is ψ after replacement of all terms with an occurrence of X free in ψ with terms that do not include X but are equivalent when X is empty. In ψ', occurrences of $X(x)$ can be replaced with $x \neq x$, and any atomic predicate expressions containing X need to be replaced with expressions that use new predicates, with fewer arguments, that handle the case of a given argument referring to an empty set (for convenience, we may wish to allow 0-ary predicates). For instance, if ψ contains the atomic formula $P(X, Y)$, we generate a new unary predicate P_1 interpreted as $P_1^{M'} = \{A : (\varnothing, A) \in P^M\}$ where M' is our model with the additional predicate. We handle names similarly to quantifiers, by replacing a sentence using the name with a disjunction of two sentences, one conditional on the name referring to the empty set and one not doing so. Furthermore, if there are numerical quantifiers, these can be fixed up to take care of the case where something is empty in a straightforward way. And if we can define counting for collections of sets that do not contain an empty set, we can define counting for all collections of sets.[15] And so the question of defining counting of complex individuals in Goodman and Quine's setting is equivalent to that of defining counting of sets (i.e., second-order objects) in MSO. Alternately, we can introduce an expressively insignificant "empty" object into Goodman and Quine's ontology. In any case, henceforth we ignore the issue of the empty set, and take MSO (with appropriate additional predicates and names as needed) as equi-expressive with atomic mereology.

Now, if Goodman and Quine can give an account of "more" for individuals, they can give an account of equinumerosity or equal count: there are equal numbers of cats and dogs if and only if there aren't more cats than dogs and there aren't more dogs than cats.

But it turns out that if we take the atomic sense of "bigger" as "having more atoms", it follows from recent work of van Benthem and Icard [11] that it is impossible to express equinumerosity of collections of complex individuals in terms of our atomic "bigger". For consider a version $\mathrm{MSO}^{>}_{\mathrm{fin}}$ of MSO with a "bigger" predicate $>$ and the restriction that the only admissible

[15]For instance, the collections A and B of sets are equinumerous if and only if either (a) neither contains the empty set and they are equinumerous or (b) one contains the empty set and nothing else while the other contains exactly one member or (c) both contain the empty set and some other members and the respective collections of these other members are equinumerous or (d) exactly one contains the empty set and at least one other member while the other does not contain the empty set and the collections resulting from removal of the empty set from the first and an arbitrary element from the second are equinumerous. This disjunction defines equinumerosity for collections that are allowed to contain the empty set in terms of equinumerosity for collections that are not allowed to contain the empty set. A similar trick can be used for "more".

models M are finite ones where $X > Y$ is interpreted as saying that the cardinality of the set denoted by X is greater than that of the set denoted by Y. We also include first-order unary predicates A and B in our language. Say that a binary relation R between natural numbers is *expressible* in $\mathrm{MSO}^{>}_{\mathrm{fin}}$ if and only if there is a sentence ψ such that for all admissible models M we have

$$M \vDash \psi \text{ if and only if } R(|A^M|, |B^M|).$$

In words, ψ holds just in case the cardinalities of the interpretations of A and B satisfy R. It follows from Van Benthem and Icard [11, Thm. 4.7][16] that a numerical relation is expressible in $\mathrm{MSO}^{>}_{\mathrm{fin}}$ if and only if it is semi-linear, from which it follows that:

Lemma 1. *There is no sentence ϕ such that for all finite models M with A^M and B^M non-empty we have $M \vDash \phi$ if and only if $|A^M| = 2^{|B^M|}$.*

The Appendix of this paper gives a definition of the semi-linearity of binary relations and details of proof given [11, Thm. 4.7].

But given Lemma 1, it follows that there is no way to define equinumerosity in MSO with $>$. For if we could define equinumerosity in MSO with $>$, we could say that the number of X such that $\forall x(X(x) \to B(x))$ (i.e., the number of subsets of the extension of B) equals the number of X such that

$$\forall x(X(x) \to A(x)) \wedge \exists x \forall y(X(y) \leftrightarrow x = y)$$

(i.e., the number of singleton subsets of the extension of A), and this would hold in M precisely when $|A^M| = 2^{|B^M|}$.

[16]See also [**?**] which corrects an error in a later part (Section 5.3) of [11]. However, the results we are relying on do not depend on that part.

Thus, there is no way to use the atomic sense of "bigger" to define equinumerosity of objects. But let us now explore some other resources that Goodman and Quine could try to draw on.

3. Extensions of Goodmand and Quine's "Steps"

3.1. **More counting predicates.** While "bigger" is the one predicate that Goodman and Quine actually used to try to count individuals, there is no reason to suppose that this is the only possible predicate that someone with their ontology could use for their purpose. Suppose for instance that Goodman and Quine had a "power of two" predicate that applies to two individuals provided that the number of atoms in the first individuals equalled two to the power of the number of atoms in the second. Then our argument that Goodman and Quine cannot compare the counts of individuals would fail.

Granted, we might be suspicious of introducing a predicate like this, given that it sounds like we are presupposing counting. But as noted in Section 2.2, such numerical relations between pairs of individuals may well be nominalistically and physicalistically acceptable. And with a bit more creativity than that involved in Trash, we can imagine a physical procedure allowing the definition of the power of two predicate with the help of a binary-counting computational machine and a rule that whether one puts an atom of y in the trash depends on what the display is showing.[17]

[17] This might require that x and y not overlap. But given that we can define having an equal number of atoms using Move or Trash, if x and y overlap, we can replace y with a y' that has the same atomic size as y but does not overlap x. Granted, this only works if the cosmos is large enough to have the extra atoms needed.

Thus, as a first attempt at helping Goodman and Quine, we might give them a larger array of predicates that compare atomic sizes of individuals than just “bigger”. To be maximally generous, thus, consider MSO with finite models, but allow the use of a predicate corresponding to every k-ary relation on the natural numbers $\mathbb{N}$ for any finite k. Given the finitism in Goodman and Quine’s “Steps” paper, they will only be willing to use finitely many of these predicates[18], but if we can show that counting comparisons cannot be expressed with *any* predicates corresponding to relations on $\mathbb{N}$, our limiting result will be all the stronger for it. Let $\mathcal{R}$ be the set of all such relations, and as usual identify such a k-ary relation with a subset of $\mathbb{N}^k$. For a k-ary relation R in $\mathcal{R}$, let C_R be the corresponding k-ary predicate, and assume that our model interprets this predicate to be satisfied by the tuple of sets $(A_1, \ldots, A_k)$ if and only if $(|A_1|, \ldots, |A_k|) \in R$, where $|A|$ is the cardinality of A. Thus, our atomically understood “bigger” predicate is just $C_>$, while the “power of two” predicate would be $C_{\{(n,m)\in\mathbb{N}^2 : n=2^m\}}$.

These cardinality predicates then correspond to atomic size predicates that go beyond “bigger”, and allow us to define any atomically-based numerical relationship between finite tuples of mereological individuals. But they are not enough to allow us to define counting of mereological individuals in general. To state the relevant MSO result, in addition to having our cardinality predicates C_R for $R \in \mathcal{R}$, add two unary predicates F and G with second-order arguments, so that our task will be to define the concept of F and G having equal size of extension. Here, however, we have an impossibility result:[19]

[18] We are grateful to an anonymous reader for this point.

[19] The result was conjectured by the first author and proved by the second.

Theorem 1. *Say that an MSO model $M = (U, \sigma, I)$ is admissible provided that its signature consists of all cardinality predicates C_R with interpretation $C_R^M = \{(A_1, \ldots, A_k) : (|A_1|, \ldots, |A_k|) \in R\}$ and the two third-order unary predicates F and G. Then there is no sentence $\phi = \phi_{F,G}$ such that for all finite admissible models M we have $M \vDash \phi$ if and only if $|F^M| = |G^M|$.*

This will be a special case of Theorem 2, below, which is proved in the Appendix.

3.2. **More predicates.** The objects to be counted are not alone in the world. The larger structure of the world may include tools that help count, like an abacus can. For instance, [9] suggested that the infinite number of points of spacetime could provide a helpful structure. However, the main results of this paper are restricted to Goodman and Quine's original setup which was explicitly meant to be compatible with a finite context.

Nonetheless, even in a finite context there are other tools available. So far the only predicates we drew on to define counting were arithmetic relations between the "atomic sizes" of mereological individuals. But there are many other physical predicates in the world, expressing properties or relations such as distance, mass, charge, shape, color, etc. Could some of these predicates help?

An initial objection to using such predicates is that objects should be able to be counted in any possible world. Now, nominalists should not quantify over properties—by the famous ontological commitment criterion of the later Quine—so any nominalistic definition of counting can only use a fixed finite number of predicates corresponding to a fixed finite number of properties. Suppose now for a *reductio* that there is a nominalistic definition of counting, and let $\mathcal{F}$ be the finite collection of predicates used in it, besides identity and our cardinality predicates. Assume, given the physicalism of Goodman and Quine's "Steps", that all these predicates are physical. Now, while in reality there might be strongly emergent properties and relations of complex individuals that cannot be defined by mereological atoms, worlds with no such irreducible predicates except perhaps for our "atomic size" predicates seem metaphysically possible. Let us assume this possibility. Then counting should work in all possible worlds, and hence also in such reductionistic worlds. So we can assume that any predicate in $\mathcal{F}$ that can be satisfied by a sequence of objects at least one of which is non-atomic can be finitely defined by predicates that are mereologically atomic in the sense that they can only be satisfied by mereological atoms. Having performed such a reduction, we can assume that $\mathcal{F}$ contains only mereologically atomic predicates.

But for any finite collection $\mathcal{F}$ of physicalistic mereologically atomic predicates other than identity, and assuming the physicalism of "Steps", it is very plausible that for any finite non-zero N there is a metaphysically possible cosmos with exactly N atoms, all of which are $\mathcal{F}$-indiscernible, i.e., such that if $F(a_1, \ldots, a_k)$ for F in $\mathcal{F}$ and distinct atoms $a_1, \ldots, a_k$, then $F(a'_1, \ldots, a'_k)$ for any atoms $a'_1, \ldots, a'_k$.[20] Just imagine a cosmos made of N bosons all of which share the same quantum state. Even someone who believes in Leibniz's Principle of the Identity of Indiscernibles can accept this because $\mathcal{F}$-indiscernibility is compatible with being discernible with respect to other

[20]The assumption of physicalism rules out cases such as deities who exist in all possible worlds and can be distinguished from all physical entities by lacking any physical attributes.

properties. But if all the atoms are $\mathcal{F}$-indiscernible in worlds of some sort, any definition of counting that uses the predicates in $\mathcal{F}$ can be modified not to use them, and will still work in worlds of the given sort—just replace any expression of the form $F(x_1, \ldots, x_k)$ for F in $\mathcal{F}$ by a tautology or the denial of a tautology, depending on whether F is satisfied by all k-tuples of atoms or by none of them (since $\mathcal{F}$-indiscernibility rules out the possibility that it is satisfied by some but not by others). And now Theorem 1 applies once again to yield a contradiction to the assumption that counting of complex individuals can be defined.

If one is suspicious of such indiscernibility assumptions, however, there is a more technical argument. Again, we can assume that the finite set $\mathcal{F}$ of predicates used in the alleged definition of counting consists entirely of mereologically atomic predicates, because there will be possible cosmoses with an arbitrary non-zero finite number of atoms where non-atomic predicates other than the atomic size predicates are definable in terms of mereologically atomic predicates. And now the possibility of defining counting for all collections of objects violates the following generalization of Theorem 1, since mereologically atomic predicates correspond to MSO's first-order predicates.

Theorem 2. *Consider a class $\mathcal{M}$ of MSO models $M = (U, \sigma, I)$ where the signature σ consists of all cardinality predicates C_R with interpretation $C_R^M = \{(A_1, \ldots, A_k) : (|A_1|, \ldots, |A_k|) \in R\}$, the two third-order unary predicates F and G, and any number of first-order predicates. Suppose that $\mathcal{M}$ contains models with universes of arbitrary non-zero finite size and that for any model $M = (U, \sigma, I)$ in $\mathcal{M}$ and any subsets H and K of the power set of U, there is a model $M_{H,K} = (U, \sigma, I_{H,K})$ where $I_{H,K}$ agrees with I on* *everything in σ other than F and G, and interprets F and G as H and K respectively. Then there is no sentence $\phi = \phi_{F,G}$ such that for all models M in $\mathcal{M}$ we have $M \vDash \phi$ if and only if $|F^M| = |G^M|$.*

The statement of Theorem 2 is somewhat complicated. The idea is that we want our definition ϕ of equinumerosity to work in models of arbitrary size, all with the same cardinality predicates C_R and some auxiliary first-order predicates like distance, mass and charge, and we want the definition to work for arbitrary collections F and G of complex individuals, which corresponds to there being models where F and G can be given any interpretation we like but that fix all the other features besides F and G. The proof of Theorem 2 is given in the Appendix.

3.3. **Objection: Asking for too much.** Now, Goodman and Quine could have objected that the above way of conceiving the task of counting is asking for too much, namely that it's asking for us to count the F-satisfiers and G-satisfiers for arbitrary interpretations. But the very statement of this task is nominalistically suspicious, since it quantifies over all interpretations of F and G. An interpretation of a unary predicate with a second-order argument is, after all, a set of subsets of the universe U or, in the mereological setting, a *set* of mereological individuals.

Instead, Goodman and Quine can insist that we should conceive the task of counting to be that of providing an account of the equinumerosity of the X's satisfying two formulas. Here is a way to make this task precise. We want to have an equinumerosity quantifier. Syntactically, this binds X in expressions of the form $\#_X\tau = \#_X\psi$ where τ and ψ have X as a free variable (and may have other free variables). Semantically, $\#_X\tau = \#_X\psi$ is supposed to be true in a model M under an assignment ρ of the remaining free variables just in case

$$|\{A : M \vDash_{\rho, A/X} \tau\}| = |\{A : M \vDash_{\rho, A/X} \psi\}|,$$

where $M \vDash_{\rho,A/X} \tau$ means that τ is true under the assignment given by ρ together with assigning A to free instances of X. The question now is whether we can find a translation of $\#_X \tau = \#_X \psi$. The language into which we perform the translation will be $\mathcal{L}$, a second-order language with the predicates C_R for $R \in \mathcal{R}$, but without F and G. Let $\mathcal{L}_{F,G}$ be the language behind Theorem 2 that does have F and G.

One way to think of the translation question is to reserve some set $\mathcal{X}$ of infinitely many first-order and infinitely many second-order variables for the free variables of τ and ψ other than X, and then to ask whether there is a schematic formula Φ_X with two types of blanks, where if we put τ in blanks of the first type and ψ in blanks of the second type we get a formula $\Phi_X(\tau, \psi)$ that holds in exactly the same models under the same assignments of the remaining free variables of $\mathcal{X}$ as $\#_X \tau = \#_X \psi$ does. The translation thesis we will now argue against is that there is such a schematic formula.

It follows from Theorem 2 that this translation thesis is false. For suppose we have such a schematic formula Φ_X. Then let ϕ be $\Phi_X(F(X), G(X))$ and by Theorem 2, let $M \in \mathcal{M}$ be a model in which exactly one of $M \vDash \phi$ and $|F^M| = |G^M|$ is true. If we could apply the translation thesis to the formula $\#_X F(X) = \#_X G(X)$, we would be done, but the translation thesis is for the language $\mathcal{L}$ that doesn't include F or G.

Instead, introduce names $a_1, \ldots, a_n$ for all the n objects in M, thereby generating a language $\mathcal{L}'$ and a model M' just like M except for these added names. Let $F'(X, a_1, \ldots, a_n)$ and $G'(X, a_1, \ldots, a_n)$ be formulas using one free second-order variable X, the names $a_1, \ldots, a_n$ and no other non-logical vocabulary, such that the assignment A/X satisfies $F'(X, a_1, \ldots, a_n)$ in M' if and only if $A \in F^M$ and satisfies $G'(X, a_1, \ldots, a_n)$ if and only if in case $A \in G^M$.[21] These formulas will offer a complete description of which sets A are in F^M and G^M in terms of their possible members $a_1, \ldots, a_n$.

Let $\mathcal{L}_0$ and $\mathcal{L}'_0$ be subsets of $\mathcal{L}$ and $\mathcal{L}'$ that use only the predicates occurring in Φ_X as well as identity. Let $\psi(a_1, \ldots, a_n)$ be a conjunction of all atomic sentences using the names $a_1, \ldots, a_n$ and first-order predicates from $\mathcal{L}'_0$ that are true in M'. Let $\psi(x_1, \ldots, x_n)$ be the formula in $\mathcal{L}_0$ resulting from replacing a_i with the first-order variable x_i for all i.

Now let ϕ_0 be the sentence:

$$\forall x_1 \ldots \forall x_n \left[\psi(x_1, \ldots, x_n) \to \#_X F'(X, x_1, \ldots, x_n) = \#_X G'(X, x_1, \ldots, x_n]\right)$$

of $\mathcal{L}_0$. Then $M \vDash \phi_0$ if and only if $|F^M| = |G^M|$: the number of satisfiers of $F'(X, a_1, \ldots, a_n)$ in M' does not depend on the identities of the a_i, just on their being distinct, and likewise for G'. Next let ϕ_1 be:

$$\forall x_1 \ldots \forall x_n \left[\psi(x_1, \ldots, x_n) \to \Phi_X(F'(X, x_1, \ldots, x_n), G'(X, x_1, \ldots, x_n)]\right).$$

But the only predicates in our schematic formula Φ_X are cardinality predicates and the predicates used in $\psi(x_1, \ldots, x_n)$, while the cardinality predicates are invariant under permutations of the universe, and $\psi(x_1, \ldots, x_n)$ includes $x_i \neq x_j$ for all i and j, so $\psi(x_1, \ldots, x_n)$ implies that the sequence $x_1, \ldots, x_n$ is a permutation of $a_1, \ldots, a_n$ and $(x_{i_1}, \ldots, x_{i_k})$ satisfies a first-order predicate used in Φ_X if and only if $(a_{i_1}, \ldots, a_{i_k})$ satisfies it. Thus,

[21] We can take $F'(X, a_1, \ldots, a_n)$ to be the disjunction of 2^n formulas of the form $\varepsilon_1 X(a_1) \wedge \cdots \wedge \varepsilon_n X(a_n)$, where ε_i is either nothing or a negation.

$M \vDash \phi_1$ if and only if $M' \vDash \Phi_X(F'(X, a_1, \ldots, a_n), G'(X, a_1, \ldots, a_n))$. But the latter holds if and only if $M' \vDash \Phi_X(F(X), G(X))$, which holds if and only if $M \vDash \Phi_X(F(X), G(X))$. Thus, $M \vDash \phi_1$ if and only if $M \vDash \Phi_X(F(X), G(X))$.

Now by our translation claim we have $M \vDash \phi_1$ if and only if $M \vDash \phi_0$, which we saw held if and only if $|F^M| = |G^M|$. But remember that exactly one of $M \vDash \Phi_X(F(X), G(X))$ and $|F^M| = |G^M|$ was true, so we have a contradiction.

4. Open questions

4.1. **Infinite worlds?** Can we do better at counting with an infinite mereologically atomic world? As a warm-up, suppose we have a "finite atomic size" predicate Fin such that $\mathrm{Fin}(x)$ for an individual x if and only if x has finitely many atoms. Then a trick of [2] shows that there is a way to express there being a finite number of mereological individuals. For note that the following two statements are equivalent in a mereologically atomic world:

(i) There are finitely many individuals x such that $F(x)$

(ii) There is an individual x of finite atomic size such that for any individuals y and z if $F(y)$ and $F(z)$, then y and z differ within x.

Here, we say that y and z *differ within* x provided that there is part of one of y and z that is not a part of the other and yet is a part of x. For if x is such as in (ii) and $F(z)$ holds, then there are at most 2^n objects y that differ from z within x, where n is the number of atoms of x. Conversely, if there are finitely many objects x such that $F(x)$, then for every pair of y and z such that $F(y)$ and $F(z)$, choose some atom a that is a part of one but not of the other, and let x_0 be an object constituted by all the chosen atoms. Then any two satisfiers of F differ within x_0.

Thus, if the task is to see if there are more individuals satisfying F than ones satisfying G, and we ignore the differences between different infinite cardinalities (it is not clear that such differences should matter to Goodman and Quine given their nominalism), the main remaining question will be whether we can account for counting comparisons of individuals composed of a finite number of atoms.

If our mereologically atomic world has enough of the right kind of structure, a positive answer can be given. For an easy case, suppose the atoms have the relational structure of the natural numbers, for instance because they are arranged along a single infinite ray, and have some physical relations between them corresponding to successor, addition and multiplication, such that the axioms of Second Order Arithmetic (SOA) hold, with mereologically complex individuals in place of sets (i.e., of second-order objects of SOA) and some tweaking to handle the lack of an empty set as in Section 2.3. But if we have SOA, and add the third-order predicates F and G guaranteed to apply to only finitely many sets, we can express that there are equal numbers of satisfiers of F and G by a method for which thanks are due to [**?**]. Any finite sequence $A_1, \dots, A_n$ of sets of numbers can be encoded as a single set of numbers, say the set $\{2^k \cdot 3^m : m \in A_k\}$, and then one can say in SOA with F that all the A_i satisfy F, that they are all distinct, and that nothing else satisfies F. Saying that there is such an encoding set then says there are exactly n satisfiers of F, from which one can define counting predicates for satisfiers of finitely-often satisfied third-order predicates.

Following [9], one might suppose space itself has the requisite structure allowing for counting to be defined for regions of space. But if we deny supersubstantialism's identification of objects with regions of space or spacetime, the world's ontology can be expected to be richer than that of points of space, because it will include individuals *in* space like electrons, dogs and cats, and so counting individuals is not the same as counting regions. Granted, if mereological atoms of individuals in space are guaranteed to have different locations, we can still count individuals by using the underlying spatial structure. However, our physics does allow for multiple bosons, such as photons, to be located in the same place in space. We can then have mereologically complex individuals constituted by pluralities of co-located photons, and it intuitively seems that the surrounding spatial structure will not help count objects that are wholly located in one place.

Nonetheless, could the move from a finite to infinite context itself help? The proof of Theorem 2 uses computational complexity considerations that no longer apply if the universe U is infinite, even if we restrict our attention to cases where F and G are satisfied by finitely many things. We thus have an open question whether Theorem 2, or at least Theorem 1, can be generalized to such contexts. Perhaps the simplest open question to state is whether Theorem 1 remains true if (a) we allow the model's universe to be infinite, (b) switch from MSO to Weak MSO where second-order variables range over finite sets, and (c) require F^M and G^M to be finite.

One may conjecture that this generalization of the theorem is still true: why should it matter for counting finite collections of finite individuals whether there are infinitely many atoms outside of them? Yet such an infinitary setting appears to significantly increase the expressive power of the language. In the finite setting, in any given formula, we can emulate quantification over natural numbers of polynomially bounded size, i.e., we can emulate formulas like $Qm(m \leq p(|U|) \to \psi(m))$ where Q is $\forall$ or $\exists$ and p is a polynomial.[22] But in the infinite setting, we can emulate quantification over all natural numbers by using the cardinalities of arbitrarily large finite sets, which seems to increase expressive power. (And then, analogously to Theorem 2, we might ask what happens if we allow other first-order predicates.)

An even more general version would be to allow the second-order variables to range over infinite sets, require the universe's cardinality U to satisfy $|U| \leq \kappa$ for some cardinality κ, and instead of allowing only cardinality predicates corresponding to subsets of Cartesian powers of $\mathbb{N}$, allow cardinality predicates that correspond to subsets of Cartesian powers of $\kappa + 1$, but still require F and G to have a finite number of satisfiers.

4.2. **A more complex translation method?** In Section 3.3, we considered a version of the task of defining equinumerosity where we have an equinumerosity quantifier ranging over mereological individuals and try to

[22]If $p(n) = n = |U|$, then this is easy: we first paraphrase $Qm(m \leq |U| \to \psi(m))$ with $QX\psi(|X|)$, and then use our cardinality quantifiers to replace $\psi(|X|)$ by something in our MSO language. But we can also handle $p(n) = (n^k - 1)/(n - 1)$ for any arbitrary finite k, and hence any polynomial bound, as follows. First paraphrase $Qm(m \leq p(|U|) \to \psi(m))$ as $Q_0 X_0 \ldots Q_{k-1} X_{k-1} \psi(|X_0| + n|X_1| + \cdots + n^{k-1}|X_{k-1}|)$. Then eliminate the constant $n = |U|$ by noting that $\forall Y(\forall x(Y(x) \leftrightarrow x = x) \to |Y| = n)$ and adding an inner quantification over Y. Then use our unlimited supply of cardinality predicates.

translate it into expressions involving atomic cardinality predicates of mereological individuals and relations between atoms. We ruled out a substitution-based translation where we have a schematic formula Φ that says that there are equal numbers of mereological individuals x such that τ as objects x such that ψ and where we can substitute any formulas open in x for τ and ψ.

But in general, translation may involve something more complicated than substitution. For instance, consider the iota operator ι, where $\iota x\phi$ is a term denoting the x such that ϕ. We cannot translate $B(\iota x K(x))$ (e.g., "The king is bald") into an expression of the form $B(\rho)$ where ρ is a term not using ι, but we can do a Russellian translation as:

$$\exists x(K(x) \wedge \forall y(K(y) \to x = y) \wedge B(x)),$$

and there is a general recursive algorithm for such rewrites of formulas with iota-expressions. However, the algorithm does not simply involve substituting subformulas using ι with subformulas not using it, since the quantifiers in the translation will in general be wide-scope.

Thus, we might ask whether there is a recursive algorithm for translating a sentence in a language $\mathcal{L}_1$ that includes atomic cardinality predicates for mereological individuals, relations between atoms and an equinumerosity quantifier into a language $\mathcal{L}_2$ that only has the atomic cardinality predicates and the relations between atoms, where the only constraint on the translation is that both sentences are true in the same atomic mereological worlds. Or, even more generally, we might ask abstractly whether for every sentence of $\mathcal{L}_1$ there is a sentence of $\mathcal{L}_2$ that is true in the same atomic mereological worlds. We do not know the answers to these questions. In the nominalist setting, the algorithmic question seems the more relevant one. Nonetheless, philosophically, we can at least say that the onus would be on the Goodman-Quine-style nominalist to show that there is such a more complex translation method. The nominalist should not simply assume there is one, especially given the negative results of this paper.

5. Conclusions

In mereologically atomic worlds, the Goodman and Quine mereological strategy for comparing the numbers of complex entities provably fails, given some assumptions about what the translation would look like.

In infinite worlds, if there is enough structure to emulate Second Order Arithmetic with the atoms behaving like numbers and supersubstantialism holds—i.e., objects are identified with regions—a version of the strategy succeeds (cf. [9]). However, supersubstantialism is a very controversial thesis. Plausibly, there are objects *in* space, and if so, then our physics gives us reason to think that there can be mereological atoms that are spatially colocated. In such a case, it is not clear whether the surrounding spatiotemporal structure helps with counting—the question depends on some open problems about the expressive power of certain monadic second-order logics.[23]

[23]We are grateful to two anonymous readers for a number of comments that have substantially improved this paper. We are also grateful to Robert Verrill for a discussion about quantum mechanics and co-location.

Appendix: Proofs

Lemma 1

A binary relation (we won't need any others) is *semi-linear* if and only if it is a finite disjunction of linear relations, and a binary relation R is *linear* provided that there exist natural numbers a, b, c and d such that $R(x, y)$ if and only if there is a natural number t such that $x = a + bt$ and $y = c + dt$.

Van Benthem and Icard [11] work with non-strict comparison $\leq$ of cardinalities while we're using the strict comparison $>$. However, each can be defined in terms of the other. Starting with $\leq$, one can define $|U| = |V|$ by the conjunction $|U| \leq |V|$ and $|V| \leq |U|$, and then define $|U| > |V|$ as $|V| \leq |U|$ but not $|U| = |V|$. Conversely, given $>$, one can define $|U| = |V|$ by saying that neither $|U| > |V|$ nor $|V| > |U|$, and then define $|U| \leq |V|$ as $|V| > |U|$ or $|U| = |V|$. Thus, whatever is expressible using $\leq$ is expressible using $>$, and vice versa. Van Benthem and Icard [11, Thm. 4.7] have shown that it is precisely the semi-linear relations that are expressible over finite models in MSO with $\leq$, and hence the same is true in $\mathrm{MSO}^{>}_{\mathrm{fin}}$.

Now the power of two relation $E_2(x, y)$ that holds whenever $x = 2^y$ is not semi-linear. For consider any semi-linear binary relation R. This will be a finite disjunction of linear relations $R_1, ..., R_n$ with R_i defined by the natural numbers a_i, b_i, c_i and d_i as in our definition of linearity. Let U be the set of all x such that $\exists y R(x, y)$ and let V be the set of all x such that $\exists y E_2(x, y)$. Then V is the set of all powers of 2. If all the b_i are zero, then $U = \{a_1, ..., a_n\}$ and so $U \neq V$. If there is a non-zero b_i, then U contains the infinite affine sequence $a_i, a_i + b_i, a_i + 2b_i, a_i + 3b_i, ...$ but the set V of powers of 2 does not contain any infinite affine sequence, so again $U \neq V$.

Thus, E_2 is not R. Hence E_2 is not semi-linear and thus there is no sentence ψ such that for all admissible models M we have

$$M \vDash \psi \text{ if and only if } |A^M| = 2^{|B^M|}. \tag{1}$$

Proof of Lemma 1. Suppose we have a sentence ϕ such that

$$M \vDash \phi \text{ if and only if } |A^M| = 2^{|B^M|}$$

for all admissible M with A^M and B^M non-empty. Then define ψ as

$$(\phi \wedge (\exists x A(x) \wedge \exists x B(x))) \vee (\neg \exists x A(x) \wedge \exists! x B(x)),$$

and we will have (1), contrary to the fact that we cannot express powers of two in $\mathrm{MSO}^{>}_{\mathrm{fin}}$. □

Theorem 2

The basic idea of *descriptive complexity theory* is that there is a tight correspondence between expressibility of properties of finite structures in various logics—such as fragments of second-order (SO) logic—and computational complexity classes. The archetypal result in this field is Fagin's theorem, which states that a class (closed under isomorphism) of finite structures is definable by an existential SO formula if and only if membership in the class is decidable in NP (nondeterministic polynomial time). More generally, a class of finite structures is SO-definable iff it belongs to the polynomial-time hierarchy PH (see [6]).

The set-up in Theorem 2 is a bit more complicated as it involves third-order predicates and the cardinality predicates C_R, but as we will show, a modification of the equivalence of SO to PH can be used to translate Theorem 2 into a question about relativized complexity classes that is easy to answer by means of standard results from complexity theory. While we formulated the theorem for MSO for consistency with the rest of the paper, the argument actually applies to full SO.

Let us first briefly introduce the complexity classes we will work with. The reader can find more details and background in Arora and Barak [1], which we will use as a basic reference; however, this textbook does not present the definitions and main results systematically relativized as needed for our purposes, so we advise the reader to follow our presentation regardless (we will often give references to [1] for unrelativized concepts with the understanding that they can be relativized in a routine manner).

Let $\{0,1\}^* = \bigcup_{n\in\mathbb{N}}\{0,1\}^n$ (where $\mathbb{N}$ is the set of nonnegative integers) denote the set of finite strings over the two-element alphabet $\{0,1\}$; we denote the length of $w \in \{0,1\}^*$ as $|w|$. The concatenation of strings u and v is written $u \smile v$. A decision problem $L \subseteq \{0,1\}^*$ is in the class P (*polynomial time*, [1, §1.6]) if there is a polynomial $p \in \mathbb{N}[x]$ and an algorithm (formally: a Turing machine [1, §1.2]) that, given any $w \in \{0,1\}^*$, decides whether $w \in L$ or not in at most $p(|w|)$ steps. Polynomial-time computability of functions $F\colon \{0,1\}^* \to \{0,1\}^*$ is defined similarly; the class of polynomial-time computable functions is denoted FP ([1, §17.2]). A problem $L \subseteq \{0,1\}^*$ is in NP (*nondeterministic polynomial time*, [1, §2.1]) if there is a polynomial $p \in \mathbb{N}[x]$ and a predicate $P \in \mathsf{P}$ such that for all $w \in \{0,1\}^*$,

$$w \in L \iff \exists u \in \{0,1\}^{p(|w|)}\, (w,u) \in P. \tag{2}$$

Here and below, (w_0, w_1) is a suitable pairing function $\{0,1\}^* \times \{0,1\}^* \to \{0,1\}^*$; we omit details of a particular definition of (w_0, w_1) as they are not important (essentially, we only need that (w_0, w_1) as well as the projections $(w_0, w_1) \mapsto w_i$ are polynomial-time computable).

We will work extensively with *relativized* complexity classes and algorithms ([1, §3.4]). An *oracle algorithm* (formally: oracle Turing machine) is an algorithm that is equipped with an *oracle* $O \subseteq \{0,1\}^*$, and—apart from normal computation steps—may ask queries of the form "is $x \in O$?" for $x \in \{0,1\}^*$, and the oracle answers each query correctly in one step. The oracle itself is not considered part of the specification of the algorithm: the same algorithm can compute different problems when it is supplied with different oracles. To stress this, we will write A^O for the instantiation of an oracle algorithm A with a specific oracle $O \subseteq \{0,1\}^*$.

A *polynomial-time* oracle algorithm is an oracle algorithm A for which there exists a polynomial $p \in \mathbb{N}[x]$ such that A^O halts in at most $p(|w|)$ steps on any input $w \in \{0,1\}^*$ with any oracle $O \subseteq \{0,1\}^*$. For a given oracle O, a decision problem $L \subseteq \{0,1\}^*$ is in the class P^O (*polynomial time relative to* O) if there is a polynomial-time oracle algorithm A such that A^O decides membership in L. A decision problem L is in NP^O if there are $p \in \mathbb{N}[x]$ and $P \in \mathsf{P}^O$ such that (2) holds. By an *oracle* NP *algorithm* A', we mean a polynomial $p \in \mathbb{N}[x]$ and a polynomial-time oracle algorithm A (without specifying the oracle), with the understanding that when supplied with any oracle O, A' decides the NP^O-problem A'^O defined by means of the expression (2), that is,

$$w \in A'^O \iff \exists u \in \{0,1\}^{p(|w|)}\, (w, u) \in A^O.$$

We will also consider algorithms (and associated complexity classes) that can query two or more oracles at the same time, say $A^{O,O'}$. We can formally define them as algorithms using the *join*

$$O \oplus O' = \{0\smile w : w \in O\} \cup \{1\smile w : w \in O'\}$$

as a single oracle.

If $\mathcal{C} \subseteq \mathcal{P}(\{0,1\}^*)$ is a class of oracles, we write $\mathsf{P}^{\mathcal{C}} = \bigcup_{O\in\mathcal{C}} \mathsf{P}^O$ and $\mathsf{NP}^{\mathcal{C}} = \bigcup_{O\in\mathcal{C}} \mathsf{NP}^O$, and similarly for other relativized classes we introduce below.

The (relativized) *polynomial hierarchy* PH^O is defined as $\bigcup_{k\geq 1}(\Sigma^{\mathsf{p}}_k)^O$, where the levels $(\Sigma^{\mathsf{p}}_k)^O$ are defined by induction on k with $(\Sigma^{\mathsf{p}}_1)^O = \mathsf{NP}^O$ and $(\Sigma^{\mathsf{p}}_{k+1})^O = \mathsf{NP}^{(\Sigma^{\mathsf{p}}_k)^O}$ ([1, §5.5]). Equivalently, $L \in (\Sigma^{\mathsf{p}}_k)^O$ iff there is a polynomial p and an oracle polynomial-time algorithm A such that for all $w \in \{0,1\}^*$,

$$w \in L \iff \exists u_1\, \forall u_2\, \dots\, Q_k u_k\, (w, u_1, \dots, u_k) \in A^O, \tag{3}$$

where Q_i is $\exists$ for odd i and $\forall$ for even i, and the quantifiers run over $u_i \in \{0,1\}^{p(|w|)}$ ([1, §5.2]). (We could also bound different u_i by different polynomials $p_i(|w|)$ instead of $p(|w|)$; this does not matter.) An *oracle* Σ^{p}_k *algorithm* A' is specified by p and A as above; for any oracle O, it decides the problem $A'^O \in (\Sigma^{\mathsf{p}}_k)^O$ defined by the right-hand side of (3).

Let $\mathcal{C}$ be any complexity class. A problem L is in the class $\mathcal{C}/\mathrm{poly}$ if there exists a predicate $L_0 \in \mathcal{C}$, a polynomial $p \in \mathbb{N}[x]$, and a *nonuniform advice* $(a_n : n \in \mathbb{N})$, where each a_n is a string of length $|a_n| \leq p(n)$, such that

$$w \in L \iff (w, a_{|w|}) \in L_0$$

([1, §6.3]). That is, we can compute L in $\mathcal{C}$ with the help of a polynomially long advice string a_n that only depends on the length n of the input, rather than on the input itself. When we speak of *algorithms* with advice, the sequence of advice strings is part of the specification of the algorithm: e.g., an *oracle* Σ^{p}_k/poly *algorithm* A' consists of an oracle Σ^{p}_k algorithm A and a sequence $(a_n : n \in \mathbb{N})$ of polynomially-long strings as above; when supplied with any oracle O, A' decides the problem $A'^O = \{w : (w, a_{|w|}) \in A^O\}$.

Finally, we need some counting classes (for which we will not bother to define a separate notion of an "algorithm"). A problem L is in the class PP^O ([1, §17.2]) if there is a $p \in \mathbb{N}[x]$ and $P \in \mathsf{P}^O$ such that

$$w \in L \iff \Pr_{u \in \{0,1\}^{p(|w|)}} \big[(w, u) \in P\big] \geq \frac{1}{2}, \tag{4}$$

where we take probability with respect to the uniform distribution over $\{0,1\}^{p(|w|)}$. A lesser-known class $\mathsf{C_{=}P}^O$ (first considered by Wagner [12]) consists of problems L that can be expressed as

$$w \in L \iff \Pr_{u \in \{0,1\}^{p(|w|)}} \big[(w, u) \in P\big] = \frac{1}{2} \tag{5}$$

with P and p as above. There are a few alternative equivalent definitions of this class: in particular, $L \in \mathsf{C_{=}P}^O$ iff there is $p \in \mathbb{N}[x]$, $P \in \mathsf{P}^O$, and $F \in \mathsf{FP}^O$ such that

$$w \in L \iff \big|\big\{u \in \{0,1\}^{p(|w|)} : (w, u) \in P\big\}\big| = F(w), \tag{6}$$

where we think of $F(w) \in \{0,1\}^*$ as representing an integer written in binary notation. This generalizes definition (5) as the function $F(w) = 2^{p(|w|)-1}$ is polynomial-time computable; in the other direction, we have

$$\big|\big\{u \in \{0,1\}^{p(|w|)} : (w, u) \in P\big\}\big| = F(w) \iff \Pr_{\alpha, u}\big[(w, \alpha, u) \in P'\big] = \frac{1}{2},$$

where the probability is over $(\alpha, u) \in \{0,1\} \times \{0,1\}^{p(|w|)} = \{0,1\}^{p(|w|)+1}$, and P' is defined by

$$(w, \alpha, u) \in P' \iff \begin{cases} (w, u) \in P, & \text{if } \alpha = 0, \\ u \geq F(w), & \text{if } \alpha = 1, \end{cases}$$

where we treat u as a binary representation of an integer $0 \leq u < 2^{p(|w|)}$ on the second line.

A decision problem $L' \subseteq \{0,1\}^*$ is *reducible* to a problem $L \subseteq \{0,1\}^*$ if there exists a function $F \in \mathsf{FP}$ such that

$$w \in L' \iff F(w) \in L$$

for all $w \in \{0,1\}^*$ ([1, §2.2]). If $\mathcal{C} \subseteq \mathcal{P}(\{0,1\}^*)$ is a complexity class, a $\mathcal{C}$*-complete problem* is an $L \in \mathcal{C}$ such that every $L' \in \mathcal{C}$ is reducible to L. Most of the uniform classes mentioned above have complete problems; in particular, we will use the fact that for each $k \geq 1$ and every oracle O, there exists a $(\Sigma^{\mathsf{p}}_k)^O$-complete problem ([1, §5.2.2]).

When we discuss decision problems where the input is a finite first-order structure M, we represent it by a finite string as follows. We assume that the domain of the structure is $[n] = \{0, \dots, n-1\}$. The representation is a tuple that includes n written in unary (i.e., as a string of 1s of length n, denoted 1^n), and the table of values of each relation and function in the signature. The table of a k-ary relation R is a string in $\{0,1\}^{n^k}$ whose $(i_{k-1}n^{k-1} + \dots + i_1 n + i_0)$th bit indicates whether $(i_0, \dots, i_{n-1}) \in R$; the table of a function can be taken to be the table of its graph (which is a $(k+1)$-ary relation). Observe that if we fix a finite signature, the size of a representation of a structure M is a fixed integer polynomial $p(n)$, where n is the size of the domain of M.

For ease of reference, we recall the statement of Theorem 2 again:

Theorem 2. *Consider a class $\mathcal{M}$ of MSO models $M = (U, \sigma, I)$ where the signature σ consists of all cardinality predicates C_R with interpretation $C_R^M = \{(A_1, \dots, A_k) : (|A_1|, \dots, |A_k|) \in R\}$, the two third-order unary predicates F and G, and any number of first-order predicates. Suppose that $\mathcal{M}$ contains models with universes of arbitrary non-zero finite size and that for any model $M = (U, \sigma, I)$ in $\mathcal{M}$ and any subsets H and K of the power set of U, there is a model $M_{H,K} = (U, \sigma, I_{H,K})$ where $I_{H,K}$ agrees with I on everything in σ other than F and G, and interprets F and G as H and K respectively. Then there is no sentence $\phi = \phi_{F,G}$ such that for all models M in $\mathcal{M}$ we have $M \vDash \phi$ if and only if $|F^M| = |G^M|$.*

As we see, the structures considered in the theorem are more complicated than plain first-order structures, and we do not want to represent the third-order predicates by listing their tables as this would make the size of the representation too large (doubly exponential in n, rather than polynomial). We treat the cardinality predicates C_R simply as an additional syntactic feature of the logic, as they have a fixed interpretation. The F and G predicates will be presented as oracles: the machine can write down the representation of a set $S \subseteq U = [n]$, which takes just n bits, and ask the oracle whether it belongs to F^M or G^M, respectively.

Lemma 2. *Let ϕ be an MSO sentence using first-order symbols, the counting predicates C_R, and third-order predicates F and G, as in the statement of Theorem 2. Then for some k, there is an oracle $\Sigma_k^{\mathsf{p}}/\mathrm{poly}$ algorithm that* *decides the set of finite models $M \models \phi$ when given the first-order part of M as input, and the F^M and G^M predicates as oracles.*

We stress that the sentence ϕ is fixed in the conclusion of the lemma; we do not claim (and it is actually false) that we can test $M \models \phi$ in $\mathsf{PH}^{F,G}/\mathrm{poly}$ when both M and ϕ are given as input.

Proof. By induction on the complexity of ϕ; for the induction, we consider also formulas with free first-order and second-order variables, the interpretation of which has to be included in the input. We may assume that ϕ is unnested (i.e., first-order function symbols can only occur in atomic subformulas of the form $f(x_1, \ldots, x_k) = y$, where x_i and y are variables), and that it contains only existential quantifiers. We consider the various possibilities for ϕ:

- A first-order or second-order atomic formula ($R(\vec{x})$, $f(\vec{x}) = y$, $x = y$, $x \in X$): this can be evaluated in P by locating the relevant entry in the table of R, f, or X.
- An atomic formula of the form $F(X)$ or $G(X)$: we ask the oracle.
- An atomic formula of the form $C_R(X_1, \ldots, X_k)$: Given the interpretations $A_1, \ldots, A_k$ of $X_1, \ldots, X_k$, respectively, we can compute $a_1 = |A_1|, \ldots, a_k = |A_k|$ in polynomial time. Observe that $a_1, \ldots, a_k \leq n$; thus, we can evaluate the formula in P/poly if we take the table of R restricted to $\{0, \ldots, n\}^k$ as nonuniform advice. This requires $(n+1)^k$ bits, which is polynomial in n (here, k is a constant determined by ϕ).
- ϕ is $\neg\phi_0$, $\phi_0 \wedge \phi_1$, or $\phi_0 \vee \phi_1$: by the induction hypothesis, ϕ_0 and ϕ_1 can be evaluated by oracle $\Sigma^{\mathsf{P}}_k/\mathrm{poly}$ algorithms for some k. Then by prenexing the quantifiers from the (3) representation, $\phi_0 \wedge \phi_1$ and

$\phi_0 \vee \phi_1$ can also be evaluated by oracle Σ^{p}_k/poly algorithms, and $\neg\phi_0$ by an oracle $\Sigma^{\mathsf{p}}_{k+1}$/poly algorithm (actually, an oracle Π^{p}_k/poly algorithm, which we didn't define). Note that we combine together the nonuniform advices of the two algorithms.

- ϕ is $\exists X\, \phi_0(X, \dots)$: by the induction hypothesis, we can test whether $(M, S, \dots) \models \phi_0(X, \dots)$ by an oracle Σ^{p}_k/poly algorithm for some k. If the domain of M is $[n]$, we have

$$(M, \dots) \models \phi(\dots) \iff \exists S \in \{0,1\}^n\, (M, S, \dots) \models \phi_0(X, \dots),$$

  which can still be tested by an oracle Σ^{p}_k/poly algorithm (using the same nonuniform advice) by combining $\exists S \in \{0,1\}^n$ with the initial quantifier in (3).
- ϕ is $\exists x\, \phi_0(x, \dots)$: similar.

□

Lemma 3. *Assume that there exists a sentence $\phi_{F,G}$ as in the statement of Theorem 2. Then there exists k such that $\mathsf{C_=P}^O \subseteq (\Sigma^{\mathsf{p}}_k)^O$/poly for all oracles O.*

Proof. Fix $\phi_{F,G}$ as in the statement. By Lemma 2, there is an oracle Σ^{p}_k/poly algorithm A for some k such that A^{F^M, G^M} decides whether $M \models \phi_{F,G}$ when given the first-order part of M as input. It follows that there is an oracle Σ^{p}_k/poly algorithm B such that $B^{H,K}$ decides whether $|H| = |K|$ for any $H, K \subseteq \mathcal{P}([n])$, taking only 1^n (a sequence of n ones) as input. To see this, we fix for every n a model $M_n \in \mathcal{M}$ with domain $[n]$. By assumption, the model $M_{n,H,K}$ that differs from M_n only by interpreting F and G as H and K (respectively) is also in $\mathcal{M}$, thus $|H| = |K|$ iff $M_{n,H,K} \models \phi_{F,G}$. This can be determined by $A^{H,K}$ when given the first-order part of M_n as input; but since this is fixed for a given n, and takes polynomially many bits to describe, we can just supply it to B as additional nonuniform advice instead of proper input.

Now, let $L \in \mathsf{C_{=}P}^O$ for some oracle O, and fix $p \in \mathbb{N}[x]$ and $P \in \mathsf{P}^O$ such that (5) holds, which can be written as

$$w \in L \iff \\ \bigl|\{u \in \{0,1\}^{p(|w|)} : (w,u) \in P\}\bigr| = \bigl|\{u \in \{0,1\}^{p(|w|)} : (w,u) \notin P\}\bigr|.$$

This condition for $w \in L$ can be tested by B with input $1^{p(|w|)}$, with oracle access to $P_w = \{u : (w,u) \in P\}$ and its complement (which are in turn computable in P^O given w as a proper input), and with nonuniform advice of length polynomial in $p(|w|)$, which is polynomial in $|w|$. All in all, we obtain $L \in (\Sigma^{\mathsf{p}}_k)^{\mathsf{P}^O}/\mathrm{poly} = (\Sigma^{\mathsf{p}}_k)^O/\mathrm{poly}$. Since L was arbitrary, $\mathsf{C_{=}P}^O \subseteq (\Sigma^{\mathsf{p}}_k)^O/\mathrm{poly}$. $\square$

We finish the proof of Theorem 2 by showing that the assumption $\mathsf{C_{=}P}^O \subseteq (\Sigma^{\mathsf{p}}_k)^O/\mathrm{poly}$ for all oracles O leads to a contradiction. This is actually the most difficult part of the proof, but fortunately for us, the work was already done by others; that is, we will only need to combine several well-known results from complexity theory.

First, we observe that any PP^O predicate as in (4) can be expressed as

$$\exists s \in \{0,1\}^{p(|w|)-1}\ \bigl|\{u \in \{0,1\}^{p(|w|)} : (w,u) \in P\}\bigr| = 2^{p(|w|)-1} + s,$$

where the predicate after the quantifier is in $\mathsf{C_{=}P}^O$ as it has the form (6). Consequently,

$$\mathsf{C_{=}P}^O \subseteq (\Sigma^{\mathsf{p}}_k)^O/\mathrm{poly} \implies \mathsf{PP}^O \subseteq (\Sigma^{\mathsf{p}}_k)^O/\mathrm{poly}.$$

Next, (the relativized version of) Toda's theorem ([10], [1, §17.4]) says

$$\mathsf{PH}^O \subseteq \mathsf{P}^{\mathsf{PP}^O},$$

whence

$$\mathsf{C_{=}P}^O \subseteq (\Sigma^{\mathsf{p}}_k)^O/\mathrm{poly} \implies \mathsf{PH}^O \subseteq \mathsf{P}^{(\Sigma^{\mathsf{p}}_k)^O/\mathrm{poly}} \subseteq \mathsf{P}^{(\Sigma^{\mathsf{p}}_k)^O}/\mathrm{poly}.$$

Finally, (the relativized version of) the Karp–Lipton theorem ([7], [1, §6.4]) says

$$\mathsf{NP}^{O'} \subseteq \mathsf{P}^{O'}/\mathrm{poly} \implies \mathsf{PH}^{O'} = (\Sigma^{\mathsf{p}}_2)^{O'}$$

for any oracle O'. We take O' to be a $(\Sigma^{\mathsf{p}}_k)^O$-complete problem: then $\mathsf{P}^{O'} = \mathsf{P}^{(\Sigma^{\mathsf{p}}_k)^O}$, and $(\Sigma^{\mathsf{p}}_l)^{O'} = (\Sigma^{\mathsf{p}}_{k+l})^O$ for any $l \geq 1$, in particular $\mathsf{NP}^{O'} = (\Sigma^{\mathsf{p}}_{k+1})^O$, $\mathsf{PH}^{O'} = \mathsf{PH}^O$, and $(\Sigma^{\mathsf{p}}_2)^{O'} = (\Sigma^{\mathsf{p}}_{k+2})^O$, whence we obtain

$$\mathsf{C_{=}P}^O \subseteq (\Sigma^{\mathsf{p}}_k)^O/\mathrm{poly} \implies \mathsf{PH}^O = (\Sigma^{\mathsf{p}}_{k+2})^O$$

for all oracles O, and hence PH^O collapses.

However, it is known that there exist oracles O such that the relativized polynomial hierarchy PH^O does not collapse, i.e., $(\Sigma^{\mathsf{p}}_n)^O \subsetneq (\Sigma^{\mathsf{p}}_{n+1})^O$ for all $n \in \mathbb{N}$ (Yao [13], Håstad [4]). In particular, $(\Sigma^{\mathsf{p}}_{k+2})^O \subsetneq (\Sigma^{\mathsf{p}}_{k+3})^O \subseteq \mathsf{PH}^O$, thus

$$\mathsf{C_{=}P}^O \nsubseteq (\Sigma^{\mathsf{p}}_k)^O/\mathrm{poly}.$$

References

[1] Sanjeev Arora and Boaz Barak, *Computational Complexity: A Modern Approach*, Cambridge University Press, 2009.

[2] Vince Bárány, Łukasz Kaiser, and Alexander Rabinovich, "Expressing cardinality quantifiers in monadic second-order logic over chains", Journal of Symbolic Logic 76 (2011), pp. 603–619.

[3] Cody S. Gilmore, "Time travel, coinciding objects, and persistence", in: Oxford Studies in Metaphysics (D. W. Zimmerman, ed.), Oxford University Press, 2007, pp. 177–198.

[4] Nelson Goodman and W. V. Quine, "Steps towards a constructive nominalism", Journal of Symbolic Logic 12 (1947), pp. 105–122.

[5] Johan Håstad, "Almost optimal lower bounds for small depth circuits", in: Randomness and Computation (S. Micali, ed.), Advances in Computing Research: A Research Annual vol. 5, JAI Press, 1989, pp. 143–170.

[6] Joel D. Hamkins, "Answer to: 'Can one say that there are equal numbers of sets satisfying formulas in Second Order Arithmetic?'", *MathOverflow* (2024). `https://mathoverflow.net/questions/483956/can-one-say-that-there-are-equal-numbers-of-sets-satisfying-formulas-in-second-o`

[7] Hud Hudson, "Simples and gunk", Philosophy Compass 2 (2007), pp. 291–302.

[8] Neil Immerman, *Descriptive Complexity*, Texts in Computer Science, Springer, 1999, xvi+268 pp.

[9] Richard M. Karp and Richard J. Lipton, "Turing machines that take advice", L'Enseignement Mathématique 28 (1982), pp. 191–209.

[10] Elselijn Kingma, "Lady parts: The metaphysics of pregnancy", Royal Institute of Philosophy Supplement 82 (2018), pp. 165–187.

[11] Laurie Paul, "A one category ontology", in: Being, Freedom, and Method: Themes From the Philosophy of Peter van Inwagen (J. A. Keller, ed.), Oxford University Press, 2017, pp. 32–62.

[12] Alexander R. Pruss, "Finite equinumerosity and FOL", Mathematics Stack-Exchange (2024). `https://math.stackexchange.com/questions/4998087/finite-equinumerosity-and-fol`

[13] Marcus Rossberg and Daniel Cohnitz, "Logical consequence for nominalists", Theoria 65 (2009), pp. 147–168.

[14] Seinosuke Toda, "PP is as hard as the polynomial-time hierarchy", SIAM Journal on Computing 20 (1991), no. 5, pp. 865–877.

[15] Johan Van Benthem and Thomas Icard, "Interleaving logic and counting", Bulletin of Symbolic Logic 29 (2023), pp. 503–587.

[16] Johan Van Benthem and Thomas Icard, "Interleaving logic and counting" (corrected version) `https://arxiv.org/pdf/2507.05219`.

[17] Klaus W. Wagner, "The complexity of combinatorial problems with succinct input representation", Acta Informatica 23 (1986), pp. 325–356.

[18] Andrew C.-C. Yao, *Separating the polynomial-time hierarchy by oracles*, in: Proceedings of the 26th Annual IEEE Symposium on Foundations of Computer Science (R. E. Tarjan, ed.), 1985, pp. 1–10.