Symmetric relations, symmetric theories, and Pythagrapheanism

Tim Button
tim.button@ucl.ac.uk

November 14, 2022

This article is forthcoming at *Philosophy and Phenomenological Research*. This is a pre-print; the published version may be subject to minor changes.

Abstract. It is a metaphysical orthodoxy that interesting non-symmetric relations cannot be reduced to symmetric ones. This orthodoxy is wrong. I show this by exploring the expressive power of symmetric theories, i.e. theories which use only symmetric predicates. Such theories are powerful enough to raise the possibility of Pythagrapheanism, i.e. the possibility that the world is just a vast, unlabelled, undirected graph.

Plenty of relations are not symmetric: I *see* Ragnar, whilst Ragnar might not see me; Ragnar *eats* breakfast, but breakfast certainly does not eat him. . . . But might every *basic* relation be symmetric?[1] The orthodox answer to this question is: *No*. More specifically, the orthodoxy holds that no interesting non-symmetric relations are reducible to symmetric ones.[2]

The imprecise caveat "interesting" is necessary. After all: suppose that some things are F and some things are not-F; suppose that $R(x, y)$ iff $F(x) \wedge \neg F(y)$; then R is not symmetric but is easily reducible. Such a relation, however, will be pretty uninteresting. The point of the orthodoxy is to deny that something as "interesting" as a total linear order could be reduced to (only) symmetric relations; to insist that, given the non-symmetric relations we actually find in the world, at least one basic relation must be non-symmetric.

My aim in this paper is to show that this orthodoxy is wrong. I will start by explaining why any attack on the orthodoxy must provide us with (what I call) *symmetric theories*, i.e. theories which use only symmetric predicates (see §1). This motivates a formal question: *how expressively powerful are symmetric theories?*[3]

[1] One might equally ask about *fundamental*, or *sparse*, or *first-class*, or *perfectly natural* relations. . . . Pick your favourite metaphysical honorific; nothing will turn on it. For the record, I harbour suspicions about the very idea of "basic" relations, but I have bracketed these suspicions in writing this paper.

[2] The orthodoxy is likely due to Russell (1903: §214), who concludes "that some asymmetrical relations must be ultimate, and that at least one such ultimate asymmetrical relation must be a component in any asymmetrical relation that may be suggested." Citing Russell's argument, MacBride (2007: 29–30) explicitly endorses the orthodoxy, and defends the orthodoxy further elsewhere (2015). Most contributors to the literature on converse relations either presuppose or assert the orthodoxy (I cite some of this sizeable literature in §8). Notable detractors from the orthodoxy are Armstrong (1997), Dipert (1997), and Dorr (2004).

[3] cf. MacBride's (2015: 191) suggestion that "there is doubtless scientific interest in establishing how far a program for paraphrasing away commitment to non-symmetric relations can extend".

Any theory can be faithfully interpreted in a symmetric theory (see §2). This provides us with a first line of attack against the orthodoxy. On its own, this may not be devastating to the orthodoxy (see §§3–4), but we can develop the point further. It turns out that a great many interesting theories are *synonymous* with symmetric theories (in the logicians' sense of synonymy; see §6). This decisively shows that the orthodoxy is wrong. Indeed, no formal barrier stands in the way of *Pythagrapheanism*: the doctrine that the world is just a vast, unlabelled, undirected graph (see §7).

To be clear, I am not a Pythagraphean. Indeed, I know of no good reason to think that every basic relation *is* symmetric (see §8). Still, contrary to received wisdom, there is no compelling reason to insist that there must be non-symmetric basic relations. For now, we should suspend judgement.

1 Properties, reduction and theorizing

In what follows, I will attack the orthodoxy. To frame my attack, it will help to begin by considering a claim which is related to the orthodoxy: *Interesting multi-place relations cannot be reduced to (one-place) properties*. Today, this is a widely-accepted status quo,[4] but it was not always thus. Let us imagine a contemporary metaphysician, Gottfried, who contests this status quo.

Gottfried aims to reduce relations to *relational properties*. He proceeds as follows. First, Gottfried insists that, for any object, a, and any two-place relation, R, we have a relational property, P^a_R. This is the property which y has iff $R(a, y)$; that is, P^a_R is $\lambda y R(a, y)$. Gottfried then claims that we can reduce any proposition which apparently involves a relation, say $R(a, b)$, to a proposition which involves only a relational property, say $P^a_R(b)$. He concludes that there is no need for basic relations.

Something is wrong with Gottfried's strategy, but it is instructive to spell out exactly what. In particular, we can learn a lot by considering this *mistaken* complaint Gottfried:[5]

> Gottfried defines P^a_R as $\lambda x R(a, y)$. So his would-be reduction of $R(a, b)$ to $P^a_R(b)$ is a "reduction" to $(\lambda y R(a, y))(b)$. And that is no reduction, but flagrantly circular: R appears in the would-be reduction!

Certainly, Gottfried's approach is circular. However, the circularity is not obviously problematic. Gottfried will insist that he has simply mentioned the to-be-reduced entities (i.e. relations) in formulating a description which picks out the more basic entities (i.e. properties). By itself, that is perfectly legitimate. To see this, consider an analogous case: picking out some unobservables as "the causes of such-and-such observable effects" would not prevent us from arguing that observables can be reduced to unobservables.

The problem with Gottfried's strategy does not, then, concern his inability to give a non-circular definition of relational properties. Indeed, Gottfried can and

[4] Thanks e.g. to Russell (1903: §§212–6) and Carnap (1928: §10).

[5] Cf. Russell's (1903: §214) complaint that "the supposed adjectives. . . presuppose R".

should shrug off any demand to define what (he thinks) is *basic*. The real problem facing Gottfried is that he cannot define what (he thinks) is *non-basic* in terms of what (he thinks) is basic; specifically, that he cannot define relations in terms of relational properties.

To get a sense of the problem, suppose that Gottfried tries to define each two-place relation, R, as the relation which x bears to y iff $P^x_R(y)$; that is, he tries to stipulate that R is $\lambda xy P^x_R(y)$. This is wholly illegitimate. Gottfried's reductionism requires that "P^x_R" be a one-place *primitive* (which picks out a *basic* property). Consequently, its superscript "x" is not a variable, but an inseparable part of a primitive expression; it is like the letter "x" in the English predicate "... *relaxes*". But once we have realized this, Gottfried's would-be definition is obviously just a bad pun.

The problem facing Gottfried is not specific to this particular would-be definition. Gottfried is running up against an elementary metalogical issue: first-order logic is decidable, but polyadic first-order logic is undecidable. This imposes a profound limitation on what Gottfried could ever hope to achieve. To make this limitation vivid, we employ a definition:

Definition 1: A theory, **T**, is *monadic* iff **T** is a first-order theory and **T**'s only non-logical primitives are one-place predicates.[6]

On Gottfried's view, only monadic theories should be used to discuss what is basic. But now consider the following result:

Proposition 2: No consistent monadic theory interprets Robinson Arithmetic, Q.[7]

Since Q is such a weak theory, it follows from Proposition 2 that Gottfried's brand of reductionism cannot even begin to handle arithmetic, or the sciences that rely upon it. This shows that Gottfried's reductionism is completely untenable, for elementary but deep metalogical reasons.

At this point, Gottfried might give up on *reduction*, and instead propose that multi-place relations are *grounded in* properties, or *supervene* on them, or some such. But I will not discuss such proposals; I am exclusively concerned with *reduction* here. Moreover, Gottfried's metaphysics is not the focus of my paper. So let me now return to that focus: the orthodoxy, according to which no interesting non-symmetric relations are reducible to symmetric ones.

The lesson I take from Gottfried's plight is simple: in order to overthrow the orthodoxy, I must (at least) show how we can theorize about non-symmetric relations using only symmetric relations. That is my overarching goal in what follows. To make that goal precise I will need a few definitions. The first is perfectly standard:[8]

[6] Henceforth, I assume without comment that all theories are first-order. This makes it harder to find symmetric theories (see Definition 4), since theories in stronger logics have greater expressive resources. Throughout, I treat = as part of the background logic; so monadic theories can use =.

[7] Throughout this paper, I use *interpretation* in the technical sense; see §A, Definition A.3.

[8] In some contexts, a metaphysically necessitated version of this definition might prove useful. However, nothing in this paper is much affected by modal considerations, primarily because my focus is on theories, which can have multiple models.

Definition 3: A relation R is *symmetric* iff both R is two-placed and $\forall x \forall y(R(x, y) \rightarrow R(y, x))$. Otherwise, R is *non-symmetric*.

Derivatively, I will say that a predicate is symmetric (in a theory) iff every interpretation (of the theory) assigns a symmetric relation to the predicate. More formally:

Definition 4: A predicate, R, is *symmetric* in $\mathbf{T}$ iff both R is two-placed and $\mathbf{T} \vdash \forall x \forall y(\mathrm{R}(x, y) \rightarrow \mathrm{R}(y, x))$; otherwise, R is *non-symmetric* in $\mathbf{T}$. The theory $\mathbf{T}$ itself is *symmetric* iff every $\mathbf{T}$-primitive is symmetric in $\mathbf{T}$.

In these terms, my plan is to attack the orthodoxy by providing symmetric theories. In fact, I will focus on providing graph theories, in this sense:

Definition 5: A theory, $\mathbf{T}$, is a *graph theory* iff $\mathbf{T}$'s only non-logical primitive is "E", which is symmetric and irreflexive in $\mathbf{T}$, i.e. $\mathbf{T} \vdash \forall x \neg E(x, x)$.

2 Faithful interpretation

My first strategy for obtaining symmetric theories is via *faithful interpretation* in graph theories. The rough idea is to code any instance of a non-symmetric relation, $R(a, b)$, by positing a uniquely describable pattern of nodes and edges which encode a "link" from a to b. Using these patterns, we can re-extract R from the graph's edge-relation, suggesting a way to attack the orthodoxy.

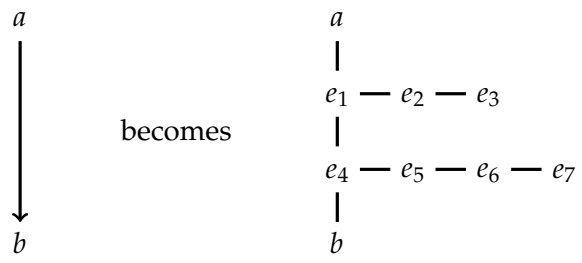
In this section, I sketch the required technicalities; I discuss the ensuing attack on the orthodoxy in §3. To be clear: the mathematics is neither mine nor new. However, whilst it is manifestly relevant to the topic of symmetric relations, it is mostly absent from the philosophical literature.[9]

To illustrate the technicalities, I will focus on a simple case. Fix a two-place non-symmetric relation, R. We can easily regard R as a directed graph, $R_{\mathbf{D}}$: its nodes are the objects in R's field, and each fact of the form that $R(a, b)$ corresponds to a directed edge from a to b.

Next, we will construct an undirected graph, $R_{\mathbf{G}}$, from $R_{\mathbf{D}}$. For each directed edge e from any a to b in $R_{\mathbf{D}}$: delete e; posit seven distinct new nodes, $e_1, \ldots e_7$;[10] and posit eight new undirected edges as follows:

[9] The mathematical idea is standard fare in model theory; see e.g. Lavrov (1963), Rabin (1965: 62), Hodges (1993: Theorem 5.5.1), and Marker (2000: 25–27). Within philosophy, Dipert (1997: 354–5) uses this sort of coding to defend his claim that the world is a graph (see also footnote 29, below).

[10] That is: nodes $e_1, \ldots, e_7$ are not in R's field, and $e_i \neq e'_j$ for each $1 \leq i, j \leq 7$ and any distinct edges e and e' of $R_{\mathbf{D}}$.



We add no other nodes or edges in constructing $R_{\mathbf{G}}$.

We now show that $R_{\mathbf{G}}$ *interprets* $R_{\mathbf{D}}$.[11] Using "E" for $R_{\mathbf{G}}$'s edge relation, define a formula which applies to all of $R_{\mathbf{G}}$'s "old" entities and none of its "new" posits:

$$Old(\mathrm{x}) :\equiv \forall v(E(\mathrm{x}, v) \rightarrow \text{(exactly 3 entities have edges to } v))$$

It is easy to check that $R_{\mathbf{G}} \vDash Old(a)$ iff a is a node in $R_{\mathbf{D}}$. Next, consider an explicit definition which we will use to simulate R:

$$\begin{aligned} R^*(\mathrm{x}, \mathrm{y}) :\equiv \text{there are } & e_1, \ldots, e_7 \text{ such that: } E(\mathrm{x}, e_1), E(\mathrm{y}, e_4), \\ & E(e_1, e_2), E(e_2, e_3), E(e_1, e_4), E(e_4, e_5), E(e_5, e_6), \text{ and } E(e_6, e_7), \\ & \text{but there are no other edges involving any of } e_1, \ldots, e_7 \end{aligned}$$

To confirm that this does indeed let us simulate R, note that for any a and b:[12]

$$R_{\mathbf{D}} \vDash R(a, b) \text{ iff } R_{\mathbf{G}} \vDash R^*(a, b)$$

With this, we have all the components we need for an interpretation of $R_{\mathbf{D}}$ in $R_{\mathbf{G}}$. Specifically, we define a translation, $*$, as follows:[13] where ϕ is a first-order formula whose only non-logical primitive is "R", let ϕ^* be the result of first restricting all of ϕ's quantifiers to "*Old*", and then replacing any subformula of the form $R(\mathrm{x}, \mathrm{y})$ with $R^*(\mathrm{x}, \mathrm{y})$. Then, for any formula $\phi(v_1, \ldots, v_n)$ with all free variables displayed, and for any "old" nodes $a_1, \ldots, a_n$:[14]

$$R_{\mathbf{D}} \vDash \phi(a_1, \ldots, a_n) \text{ iff } R_{\mathbf{G}} \vDash \phi^*(a_1, \ldots, a_n)$$

This suggests a method for reducing the non-symmetric relation, R, to a symmetric relation: claim that $R_{\mathbf{G}}$'s edge relation, E, is more basic than R, and that R is perspicuously analysed via R^*.

[11] For a formal definition of what it means for one structure to interpret another, see e.g. Hodges (1993: §5.3) and Button and Walsh (2018: §5.3).

[12] *Proof.* Left-to-right is immediate. For right-to-left, let $e_1, \ldots, e_7$ witness that $R^*(a, b)$. Inspecting the pattern of edges, all of $e_1, \ldots, e_7$ are *new*. Since $E(a, e_1)$, $E(b, e_4)$ and $E(e_1, e_4)$, the nodes $e_1, \ldots, e_7$ were introduced in the course of replacing a single directed edge between a and b; and the pattern of (undirected) edges indicates that the original edge ran from a to b, i.e. that $R(a, b)$.

There is no risk of a significant use/mention confusion here, but logical hygiene compels me to note that I am treating the structure $R_{\mathbf{D}}$, which was generated by the relation, R, as a structure whose signature involves exactly one non-logical primitive, "R". Throughout this paper, I tend to rely on context to distinguish between use and mention, explicitly marking mentions only when it seems likely to aid clarity.

[13] As in §A, Definition A.1, this is an identity-preserving translation.

[14] *Proof.* Induction on complexity, with the base case given in footnote 12.

Of course, much more would need to be said to make that claim metaphysically plausible. As MacBride puts it: "it's never enough to have some equivalence before us – the privilege has to be earned to read the equivalence as an analysis, assigning priority to one side rather than another."[15] Still, before we even start to consider how to earn that privilege, we must confirm that this method would at least meet the condition laid down in §1; i.e. that, in principle, applying this method would leave us able to theorize about non-symmetric relations.

To show that it does, suppose that $\mathbf{T}$'s only primitive is "R". Then, with $*$ defined as above, let $\mathbf{T}_{\mathsf{new}}$ be the graph theory whose axioms are exactly ϕ^*, for any $\mathbf{T}$-axiom ϕ, plus an extra axiom which ensures that $\mathbf{T}_{\mathsf{new}}$ is a graph theory, i.e. "E is symmetric and irreflexive". It is now easy to show that $*$ is a *faithful interpretation*, in that:[16]

$$\mathbf{T} \vdash \phi \text{ iff } \mathbf{T}_{\mathsf{new}} \vdash \phi^*, \text{ for any } \mathbf{T}\text{-sentence } \phi$$

This shows that $\mathbf{T}$ can be given a symmetric underpinning, in the form of $\mathbf{T}_{\mathsf{new}}$.

I have illustrated the simplest case of interpretation, where we are dealing with a single two-place relation. This is indeed artificially simple; ultimately, we will want to consider multiple different relations, with any number of places. However, with no great ingenuity but a bit of elbow grease, the strategy can easily be extended. By this means, we can obtain the following general result:

Proposition 6: Let $\mathbf{T}$ be a first-order theory, with only countably many non-logical primitives. Then some graph theory faithfully interprets $\mathbf{T}$.[17]

3 Attacking and defending the orthodoxy

I just surveyed a few results concerning (faithful) interpretation. These results are significant, given my purposes, because they allow us to formulate a crisp attack on the orthodoxy:

> The orthodoxy is that no interesting non-symmetric relations can be reduced to symmetric ones. Whilst "interesting" is imprecise, the discussion of §2 shows that *all* relations—"interesting" or not—are in principle reducible to symmetric relations. Admittedly, the reductions suggested by §2 might be metaphysically implausible, but that is neither here nor there; the point is just that reductions are *always* in principle available. And that is just what the orthodoxy denied.

This attack has considerable merit. However, as I will now explain, the orthodoxy has been formulated with enough slack to allow it to wriggle free.

---

[15] MacBride (2015: 193).

[16] *Proof.* Fix any relation R and formula ϕ. Then $R_{\mathbf{D}} \vDash \mathbf{T} \cup \{\neg\phi\}$ iff $R_{\mathbf{G}} \vDash \mathbf{T}_{\mathsf{new}} \cup \{\neg\phi^*\}$, by the earlier biconditional. Hence $\mathbf{T} \nvdash \phi$ iff $\mathbf{T} \cup \{\neg\phi\}$ is satisfiable iff $\mathbf{T}_{\mathsf{new}} \cup \{\neg\phi^*\}$ is satisfiable iff $\mathbf{T}_{\mathsf{new}} \nvdash \phi^*$. Now use soundness and completeness.

[17] Specifically, the interpretation will be identity-preserving (see §A, Definition A.3). Proposition 6 will have been known since at least the mid-1960s; I have not found it stated in this form, but the references mentioned in footnote 9 (or the discussion of §C) provide all the tools required to prove it.

Consider the total linear order, $<$, on the natural numbers. If we want to use the strategy of §2 to account for the particular arithmetical fact that $1 < 2$, for example, then we will posit seven "new" entities, $e_1, \ldots, e_7$, along with eight undirected edges. Since $e_1, \ldots, e_7$ are "new" entities, they are not themselves natural numbers. So they lie outside the field of $<$ (under interpretation). And such considerations suggest a rebuttal, on behalf of the orthodoxy, against the earlier attack:

> A paradigm of an interesting non-symmetric relation is a *total* linear order, i.e., a linear order whose field is *unrestricted*. You have posited "new" entities, and then described a relation whose field is only the "old" entities. So you have given us a relation with a *restricted* field, when precisely what we wanted was un *unrestricted* relation. Your proposed reduction of a total linear order is not, then, even adequate in principle.

This rebuttal can be glossed either as a clarification of which relations are "interesting", or as a clarification of what "adequate reduction" requires (in this context). Either way, the clarification is not unreasonable, and I will not contest it.[18] Instead, I will simply aim to make the clarification of the orthodoxy more precise (see §4), so that I can then refute it (see §6).

4 Synonymy

In this section, I will explain how we can and why we should articulate the (clarified) orthodoxy in terms of *synonymy*.

Note that here—and throughout this paper—I am not using *synonymy* in the ordinary-language-sense, according to which "eft" and "juvenile newt" are synonymous. Instead, I am speaking about an explicitly-defined relation, which holds between formal theories and is studied within mathematical logic. I should start by outlining this relation (for further detail, see §A).

Roughly speaking, to say that two theories, $\mathbf{T}$ and $\mathbf{S}$, are synonymous is to say that each interprets the other, and that combining the interpretations gets us back exactly where we began. More precisely, it is to say that there are interpretations, $\sharp : \mathbf{T} \longrightarrow \mathbf{S}$ and $\flat : \mathbf{S} \longrightarrow \mathbf{T}$, such that

(1) $\mathbf{T} \vdash \phi \leftrightarrow \phi^{\sharp\flat}$, for any $\mathbf{T}$-formula ϕ; and
(2) $\mathbf{S} \vdash \phi \leftrightarrow \phi^{\flat\sharp}$, for any $\mathbf{S}$-formula ϕ.

In this context, there is a particular interest in interpretations which combine to "get us back exactly where we began".[19] Recall that I am considering the metaphysical project of attempting to reduce *all* relations to a certain "starter-pack" of *basic* relations. Idealizing rather a lot: suppose that $\mathbf{B}$ is a total theory of what is basic,

[18] This is partly for dialectical reasons; as Lakatos's (1976) famous discussion of "monster-barring" shows, it is very difficult to police "clarifications".

[19] NB: *in this context*. In §2 I discussed (faithful) interpretation; now I am discussing synonymy; but there are many well-defined, intermediate notions, such as mutual (faithful) interpretability and bi-interpretability (see §A, Definitions A.3 and A.5) which are of great interest in other contexts.

expressed solely in terms of what is basic,[20] whilst **E** is a total theory of everything, basic and reducible alike, expressed in more compendious terms. If the intended reduction really is successful, then moving back and forth between **B** and **E** will indeed get us back "exactly where we began".[21]

Of course, synonymous theories may not be equally easy to use. But synonymy allows us to transfer work done in one theory over to the other, without loss (or gain).[22] So, if we can show that our hypothetical theories **B** and **E** are synonymous, then we can continue to *use* **E** for convenience, whilst insisting in good faith that **B** accurately reflects the *basic* state of things. This is exactly as we would hope, given the intended (reductionist) metaphysical project.

I have explained and discussed synonymy in syntactic terms. To fix the idea more firmly, though, it might help to provide a semantic gloss of synonymy. Very roughly, the thought is that models of synonymous theories differ only in their choice of (primitive) notation. More precisely, where $*$ is a translation into $\mathcal{M}$'s signature, $*\mathcal{M}$ is the structure we obtain by deploying $*$ within $\mathcal{M}$ (see §A, Definition A.2). Then, where $\sharp$ and $\flat$ are the interpretations witnessing that **T** and **S** are synonymous, we have:

(1) if $\mathcal{M} \vDash \mathbf{T}$, then $\flat\mathcal{M} \vDash \mathbf{S}$ and $\mathcal{M} = \sharp\flat\mathcal{M}$; and
(2) if $\mathcal{M} \vDash \mathbf{S}$, then $\sharp\mathcal{M} \vDash \mathbf{T}$ and $\mathcal{M} = \flat\sharp\mathcal{M}$.

Equipped with the idea of synonymy, let us revisit the discussion of §2. We saw that any theory is *faithfully* interpreted by some symmetric theory; indeed, by some graph theory. But the original theory and the graph theory may not be *synonymous*. For a simple example, let TED be the complete first-order theory of this two-element directed graph:

$$\bullet \longrightarrow \bullet$$

Every model of TED has exactly two elements; every model of $\mathsf{TED}_{\mathsf{new}}$ has more than two elements; so $\mathsf{TED}_{\mathsf{new}}$ is not synonymous with TED.

This sheds light on the attack and rebuttal I discussed in §3. We may start with a theory, **T**, which describes an apparently unrestricted relation, R; but precisely because **T** and $\mathbf{T}_{\mathsf{new}}$ not synonymous, $\mathbf{T}_{\mathsf{new}}$ interprets R as a *restricted* relation, R^*. And this suggests suggests that defenders of the orthodoxy should clarify (or reformulate) their doctrine as follows: *no interesting theory is synonymous with any symmetric theory*.[23] Or, with greater brevity: *every interesting theory is unsymmetrizable*, where we stipulate:

[20] Perhaps **B** is the Book of the World, in Sider's (2011) sense.

[21] For those already familiar with the notion of synonymy, I can make the point another way: if the intended reduction is successful, then **E** should be a *definitional extension* of **B** (see e.g. Hodges 1993: 59–61).

[22] Interpretations which witness a synonymy are always *faithful* (see §A, Definition A.3).

[23] The idea here is to clarify the orthodoxy by considering some equivalence relation between theories (specifically, *synonymy*). Now, synonymy is the tightest notion of equivalence between theories (short of syntactic identity) which is routinely studied by logicians. So, by clarifying the orthodoxy in terms of synonymy, rather than some looser equivalence relation, I am making my task (of attacking the clarified orthodoxy) as difficult as possible.

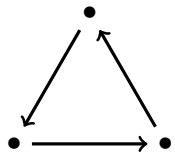
Definition 7: A theory, **T**, is *unsymmetrizable* iff no symmetric theory is synonymous with **T**.

Of course, the clarified orthodoxy still uses the caveat "interesting". But the claim is precise enough to investigate. Indeed, it is precise enough to refute.

5 Digression: unsymmetrizable theories

My refutation of the orthodoxy will come in §6. Before that, I want to make a mitigated concession: in its present formulation, the orthodoxy does contain a grain of truth. Specifically: *some* interesting theories are unsymmetrizable.

One of the simplest unsymmetrizable theories is the theory of a three-element cycle, i.e., the complete theory of this directed graph:



Admittedly, the theory of a three-element cycle is not especially interesting. However, the following result provides us with some more interesting examples of unsymmetrizable theories (for proofs, see §B):

Proposition 8: Each of these theories is unsymmetrizable:

(1) for any $n \geq 3$, the first-order theory of an n-element cycle;
(2) the first-order theory of $(\mathbb{Q}, <)$, i.e. the rationals in their usual order;
(3) the first-order theory of $(\mathbb{Z}, <)$, i.e. the integers in their usual order;
(4) Robinson Arithmetic, Q.

This illustrates the expressive limitations of symmetric theorizing. Indeed, it is especially striking that, given Propositions 2 and 8(4) respectively, if we use only monadic or symmetric predicates, then no amount of ingenuity will ever get us to (anything synonymous with) Q.

That said, there is a crucial difference between Propositions 2 and 8. Propositions 2 tells us that *every* consistent theory which interprets Q is not synonymous with any monadic theory; that was sufficient to destroy Gottfried's metaphysical ambitions. By contrast, Proposition 8(4) tells us only about Q *itself*, not about *every* consistent theory which interprets Q. Indeed, Proposition 8 allows that theories which are stronger than Q—theories with greater definitional resources—might be synonymous with some symmetric theory. So Proposition 8 is insufficient, on its own, to save the orthodoxy.

Let me put this point more plainly. Sure: Proposition 8 shows us that *some* interesting theories are unsymmetrizable. But the orthodoxy says that *all* interesting theories are unsymmetrizable. And, as I will now show, that universal claim is false.

6 Graphable theories

Some of our most important mathematical theories are *not* unsymmetrizable. Indeed, they are so far from being unsymmetrizable, that they can be rewritten as graph theories. Specifically, consider another definition:

Definition 9: A theory, **T**, is *graphable* iff **T** is synonymous with some graph theory.

My refutation of the orthodoxy really comes down to this point: *Vast swathes of mathematical theories are graphable.*[24] To show this, I will invoke the following result:

Proposition 10: Let **T** be a first-order theory, with finitely many primitives, which directly interprets $\mathsf{AS_e}$. Then **T** is graphable.

I will prove Proposition 10 in §C (indeed, I will prove a somewhat stronger result, the Graphability Theorem). Here, I will just explain what Proposition 10 means.

Proposition 10 mentions $\mathsf{AS_e}$. This is Adjunctive Set theory with extensionality.[25] This theory uses two primitives, $\in$ and *Set*, and has just three axioms:

$$\exists a(Set(a) \wedge \forall x\ x \notin a)$$
$$\forall a \forall b \exists c(Set(c) \wedge \forall x(x \in c \leftrightarrow (x \in a \vee x = b)))$$
$$\forall a \forall b((Set(a) \wedge Set(b) \wedge \forall x(x \in a \leftrightarrow x \in b)) \rightarrow a = b)$$

The first axiom says that there is an empty set; the second says that for any a and b, the set $a \cup \{b\}$ always exists; the third says that sets are extensional. Clearly, $\mathsf{AS_e}$ is a very minimal theory of sets.

Proposition 10 also mentions *direct interpretation*. A direct interpretation is an interpretation which acts only on the interpreted theory's non-logical primitives.[26] So, to say that **T** directly interprets $\mathsf{AS_e}$ is to say that there are definable **T**-formulas, $Set^*(x)$ and $x \in^* y$, and that **T** proves each of $\mathsf{AS_e}$'s three axioms, if we systematically replace *Set* with Set^* and $\in$ with $\in^*$ throughout each axiom.

The upshot of all this is that the hypothesis of Proposition 10 is *very easy to meet*. After all, $\mathsf{AS_e}$ is extremely weak, so that it is very easy to directly interpret $\mathsf{AS_e}$. Moreover, any extension of a theory which directly interprets $\mathsf{AS_e}$ also itself directly interprets $\mathsf{AS_e}$. So plenty of theories meet the hypothesis of Proposition 10. For example, Proposition 10 immediately entails that each of these theories is graphable:[27]

Kripke–Platek set theory, with or without urelements;

[24] Given the titles of their papers, one might expect this to follow from the work of Quine (1954) and Cobham (1956). However, as Visser (2008: 313) notes "It is not very clear what precisely Quine proves"—consider e.g. Quine's quantification into subscript position (1954: 180(1)). Tarski (1954: Theorem II) announces a result which seems more precise than Quine's.

[25] $\mathsf{AS_e}$ was introduced by Szmielew and Tarski (1950) as system $\mathfrak{F}'$, and discussed by Tarski et al. (1953: 34) as system S. Thanks to Allen Hazen for alerting me to the history, and to Albert Visser for suggesting the name "$\mathsf{AS_e}$".

[26] See §A, Definition A.3.

[27] The set theories on this list interpret $\mathsf{AS_e}$ verbatim. PA directly interprets $\mathsf{AS_e}$ using well-known coding tricks (see e.g. Kaye and Wong 2007).

Zermelo set theory, with or without urelements;
New Foundations, with or without urelements;
first-order Peano Arithmetic, PA.

These are all paradigms of interesting theories, describing interesting relations.

The consequences of Proposition 10 do not end there. Take any theory on the list just given, or just start with $\mathsf{AS_e}$ itself. Next, enrich your chosen theory with some first-order axioms—as many as you like. If you want, you may formulate these axioms using new primitives, provided that you use only finitely many new primitives. Now: whatever you did, the resulting theory is graphable. In a precise sense: *you could have done it all with just one symmetric relation*.

Granted, not every theory is graphable. There are things we cannot do using a single symmetric predicate. Indeed, as Proposition 8 shows, there are things that we cannot do using *any number* of symmetric predicates. Still, it is hard to imagine how there could be a more damning and complete refutation of the orthodoxy than Proposition 10 and its corollaries. For they show how vast swathes of mathematics are (perfectly) reducible to graph theories.

7 Pythagrapheanism

To ram home the failure of the orthodoxy, I will explain how it allows us to entertain the possibility of *Pythagrapheanism*.

Quine once floated the possibility of hyper-Pythagoreanism: the view that every single thing is a pure set. Briefly put, Quine's line of thought ran as follows: physics can be treated in terms of assignments of real-numbered values to spacetime regions; spacetime regions can be reduced to sets of spacetime points; spacetime points can be reduced to quadruples of real numbers; and real numbers can be reduced to pure sets in some canonical fashion.[28]

Assume for now that Quine was right, and that hyper-Pythagoreanism should be a live possibility (in some sense). Then the hyper-Pythagorean's theory of pure sets—which the hyper-Pythagorean regards as a theory of absolutely everything—will surely be an extension of $\mathsf{AS_e}$, since $\mathsf{AS_e}$ is so weak. So, by Proposition 10, their theory of pure sets is graphable. We can therefore push the hyper-Pythagorean's reduction one step further, reducing their world of sets to an undirected graph. That is: we can suggest that *everything* is (fundamentally) a node in a graph, and that the *only* basic relation is the edge relation of that graph. This is the doctrine I call *Pythagrapheanism*.[29]

In fact, this route to Pythagrapheanism sells Proposition 10 short; we can do better than piggy-backing on Quine. To explain: Quine's route to hyper-Pythagoreanism essentially amounts to explaining how to *embed* one structure ("the

[28] Quine (1976: 499–503).

[29] Dipert (1997) advances something very much like Pythagrapheanism, but with a few differences. *First*: he does not invoke Proposition 10, but relies upon the kind of coding discussed in §2 (1997: 354–5, see footnote 9). *Second*: he suggests that worldly entities are not the nodes of the graph, but its subgraphs (1997: 352-6). *Third*: he insists (1997: 348ff) that the graph should have no non-trivial automorphisms. *Fourth*: he (1997: 352, 355) is almost exclusively focussed on finite graphs.

physical world") within another ("the set hierarchy"). Now, one might well deny that hyper-Pythagoreanism is correct.[30] But we should first ask whether Quine's route even yields a *reduction*. It is one thing to show how to embed the physical world within the hierarchy of pure sets. It is quite another thing to explain how physicists might *theorize* in purely set-theoretic terms. Quine has not done this second thing: he has not explained how, given a physical theory, we can obtain a set theory from which we can "recover" the physical theory without gain or loss. Moreover, as Quine himself admits,[31] it is not immediately obvious that this second thing is even *possible*. (To see there is a real issue here, recall §1: Gottfried showed us how we can *embed* a world of relations into a world of relational properties, but he demonstrably *could not* tell us how to theorize using only monadic predicates.)

Fortunately, given two fairly minimal assumptions, Proposition 10 allows us to bypass these worries. The first assumption is that our favourite physical theory can be formulated so that it directly interprets $\mathsf{AS_e}$; an easy way for this to happen is if physicists are willing to allow their objects to be members of sets (in the sense of $\mathsf{AS_e}$).[32] The second assumption is that our favourite physical theory uses only finitely many non-logical primitives; this is scarcely an assumption at all, given the kinds of theories we actually *use* (rather than abstractly *describe*). Given both assumptions, though, our favourite physical theory is graphable, by Proposition 10. The Pythagraphean possibility now looms into view, complete with an account of how to theorize *as* a Pythagraphean: just use the synonymous graph theory in place of the original physical theory.

In sum: there is no formal impediment to the claim that you, me, and everyone we know are all just nodes in an enormous graph, and that all the various non-symmetric relations—Love, Hate, and everything else—reduce to that graph's edge relation.[33] Otherwise put: the orthodoxy is *so* wrong, that perhaps every relation reduces to a *single*, *symmetric* relation.[34]

---

[30] Kemp (2017) provides an extensive discussion and defence of hyper-Pythagoreanism; much of this could easily be tweaked to provide a defence of Pythagrapheanism.

[31] Quine (1976: 503); hence he suggests that he has reduced the "ontology" but not the "ideology".

[32] As mentioned earlier: I obtain Proposition 10 from a slightly stronger result (the Graphability Theorem of §C); so we could weaken the directly interpreted theory from $\mathsf{AS_e}$ to $\mathsf{AS_p}$ (for all this, see §C, especially footnote 53). Significantly, though, the Graphability Theorem only applies to theories without finite models. So if our favourite physical theory admits finite models, I have no general proof that it is graphable. There is much more to say about the issue of finite models; I make a small start on this in §D.

[33] In fact, **T**'s graphability can be witnessed by a theory which states that no nodes are more than two steps away from each other (see §C, Corollary C.9). This would rather trivialize the game *Six degrees of separation*.

[34] NB: Pythagrapheans do not need to deny that there *are* people, non-symmetrically loving/hating each other; they are only making a claim about what is *basic*. To vary the metaphysical honorific (see footnote 1): consider a Pythagraphean who has been inspired by Sider's (2011) project of writing the Book of the World. Our Siderean-Pythagraphean will agree that we successfully express many truths using non-symmetric predicates. Their point is that such predicates are not *joint-carving*; for a joint-carving predicate would have to pick out a fundamental relation, and (they claim) there is only one such relation, the symmetric and irreflexive Edge relation.

8 Epistemic possibility and metaphysical actuality

The "perhaps", in the last paragraph, flags an epistemic possibility. This epistemic possibility is made available by a purely formal result (Proposition 10). However, at the risk of stating the obvious: epistemic possibility does not entail metaphysical possibility, let alone actuality. So let me be clear. I am not *endorsing* the doctrine of Pythagrapheanism. Indeed, I am not endorsing the weaker claim, that every basic relation is symmetric. I am simply calling attention to an intriguing metaphysics, which—due to the orthodoxy—has been almost wholly overlooked.

Indeed, it is worth emphasising the extent of the gap between epistemic possibility and metaphysical possibility. To do this, I will consider what I believe to be the best argument to the conclusion that every basic relation is symmetric, and see how this argument fares in the light of my earlier discussion. The argument in question uses just these three premises:

DISTINCTNESS. If R is not symmetric, then $R \neq \breve{R}$ (where $\breve{R}$ is R's converse).[35]

REASONS. If R is basic and S is not, then there should be some sufficient reason for why this is so.[36]

AUSTERITY. There is no redundancy among the basic relations; the world can only be completely characterized by mentioning every basic relation.[37]

Note that I do not want to *endorse* this argument, so I will offer no defence of these premises. However, none of them is wildly implausible. And they jointly entail that every basic relation is symmetric. To see this, suppose for reductio that R is basic but not symmetric. Then $\breve{R}$ is distinct from R, by DISTINCTNESS. Furthermore, by REASONS, $\breve{R}$ is basic: after all, literally anything we could do with R, we could do with $\breve{R}$ instead, so there cannot be a non-arbitrary explanation of *why* R but not $\breve{R}$ is basic.[38] So both R and $\breve{R}$ are basic.[39] But anything characterized in terms of both

[35] This is pretty plausible. After all, if R is not symmetric, then there are a and b such that $R(a, b)$ and $\neg\breve{R}(a, b)$. However, the (apparently) obvious inference to $R \neq \breve{R}$ has been contested; see e.g. Cross (2002: 220–3), Dixon (2018), Fine (2000: 10–32), MacBride (2014: 3–4), and Williamson (1985: 256–62).

[36] Amijee (2020) provides a nice discussion of contemporary commitment to principles like REASONS.

[37] Lewis (1983: 346, 1986: 60) explicitly endorses this; see Sider (2011: 217–22, 2020: 107–10) for critical exploration.

[38] Liebesman (2014: 411) and MacBride (2007: 26, 2014: 9–10, 2020: §4) present similar considerations concerning REASONS.

[39] Plausibly, it is metaphysically necessary that $\forall x \forall y (R(x, y) \leftrightarrow \breve{R}(y, x))$. So, if both R and $\breve{R}$ are basic, then there is a metaphysically necessary connexion between basic entities. This suggests an alternative argument to the conclusion that all basic relations are symmetric, which departs from the argument in the main text by invoking this principle in place of AUSTERITY:

RECOMBINATION. There are no metaphysically necessary connexions between distinct basic relations.

Armstrong (1997: 90–91, 143–5) essentially presents this alternative argument; Dorr (2004: 161ff) presents a more complicated argument, but using a principle which essentially amounts to RECOMBINATION (his "POSSIBILITY").

In the main text, I focus solely on AUSTERITY. This involves no loss of generality, because RECOMBINATION straightforwardly entails AUSTERITY. To see this, suppose that AUSTERITY fails. So there is redundancy among what is basic. Let S be basic and redundant; so the world can be completely characterized without mentioning S, i.e. any S-involving fact is (metaphysically) determined by completely specifying the behaviour of every basic entity except S. Hence there are necessary connexions between S and the other basic relations, contradicting RECOMBINATION.

R and $\check{R}$ could equally have been characterized (exclusively) in terms of R. So there is redundancy among the basic entities, contradicting AUSTERITY. Discharging the reductio: every basic relation is symmetric.

At the risk of repetition: I do not *endorse* this argument. But I do take it seriously, and want to explore how it fares in the light of §§2–7. To make the exploration more vivid, I will introduce a character, Bella, who insists that all basic relations are symmetric because she endorses DISTINCTNESS, REASONS and AUSTERITY.

Bella needs the orthodoxy to fail. Since the discussion of §§2–7 shows the orthodoxy to be wrong, it certainly gives her *some* cause to celebrate. However, Bella's celebrations should be limited. The orthodoxy is only one potential barrier confronting her, and her adherence to both REASONS and AUSTERITY raises another considerable barrier.

To appreciate that barrier, let $\mathbf{T}$ be a theory which Bella initially regards as an excellent candidate for being *the* fundamental theory (i.e. the theory whose primitive predicates correspond bijectively with the basic relations). However, on closer inspection, $\mathbf{T}$ turns out not to be symmetric. Disappointed, Bella is forced to deny that $\mathbf{T}$ is *the* fundamental theory. Unwilling, though, to give up entirely on $\mathbf{T}$'s promise, she hits upon a plan. Assuming—modestly—that $\mathbf{T}$ meets the hypothesis of Proposition 10, there is some symmetric theory, $\mathbf{S}$, which is synonymous with $\mathbf{T}$. Since $\mathbf{S}$ is symmetric, Bella suggests that $\mathbf{S}$, rather than $\mathbf{T}$, is *the* fundamental theory.

Alas, Bella is moving much too quickly. Whilst $\mathbf{S}$ is symmetric, it may be unacceptable to Bella on other grounds. Indeed, if $\mathbf{S}$ is obtained via my strategy for proving Proposition 10, then $\mathbf{S}$ will be *exactly* as unacceptable to Bella as $\mathbf{T}$ itself. But this point will take some explaining.

To prove Proposition 10, I describe a mechanism which, given an input theory, $\mathbf{T}$, explicitly constructs a synonymous graph theory, $\mathbf{T}_{\text{graph}}$. Crucially, some of $\mathbf{T}_{\text{graph}}$'s specifics are wholly arbitrary; equally good, alternative, graph theories witness that $\mathbf{T}$ is graphable. To see this in detail, we would need to consider my mechanism for constructing $\mathbf{T}_{\text{graph}}$ from $\mathbf{T}$. Fortunately, we can leave the full, gory details to §C; here, it suffices to note that the mechanism involves several *coding choices*. These are choices like those I made, in §2, to construct $R_{\mathbf{G}}$ from $R_{\mathbf{D}}$; and, just as in §2, umpteen other choices would have worked equally well. (Indeed, even enumerating $\mathbf{T}$'s primitive predicates in a different order will yield an equally good alternative to $\mathbf{T}_{\text{graph}}$.)

Bearing all this in mind: let $\mathbf{T}_{\text{alt}}$ be one of these equally good, alternative, graph theories. Since $\mathbf{T}_{\text{graph}}$ and $\mathbf{T}_{\text{alt}}$ are *alternative* graph theories, they posit distinct edge relations; let these be E_{graph} and E_{alt}, respectively. Since $\mathbf{T}_{\text{graph}}$ and $\mathbf{T}_{\text{alt}}$ are synonymous, anything characterizable in terms of both E_{graph} and E_{alt} can be characterised in terms of just one. By AUSTERITY, then, the two relations cannot *both* be basic. But, since $\mathbf{T}_{\text{graph}}$ and $\mathbf{T}_{\text{alt}}$ are *equally good* alternatives, there cannot be a sufficient reason for why E_{graph} is basic, rather than E_{alt}. By REASONS, then, E_{graph} is not basic.

Summarizing: given both REASONS and AUSTERITY, $\mathbf{T}_{\text{graph}}$ cannot be *the* fundamental theory. Now, Bella endorses both REASONS and AUSTERITY; indeed, this is why she thinks that all basic relations are symmetric, and so spurns $\mathbf{T}$. Bella must therefore spurn $\mathbf{T}_{\text{graph}}$ too.

Here is the more general moral. Given a theory, **T**, meeting modest assumptions, Proposition 10 shows that **T** is synonymous with some symmetric theory (indeed, with some graph theory). It does *not* show that **T** is synonymous with some symmetric theory which could ever be regarded as *the* fundamental theory (by anyone who, like Bella, has even remotely sensible motivations for insisting that all basic relations are symmetric).[40] And all of this emphasises the vast gap, between establishing that many interesting non-symmetric relations are *formally* reducible to symmetric ones—as Proposition 10 does—and establishing the metaphysical thesis that every basic relation *is* symmetric.

9 Conclusion

I have not given any reason to think that every basic relation is symmetric. Still, I have shown that the orthodoxy is wrong, and spectacularly so. We cannot, yet, foreclose the possibility that every basic relation is symmetric. Indeed, we cannot even foreclose the Pythagraphean possibility, according to which there is only one basic relation, and a symmetric one at that. The technical results discussed in this paper provide a proof of concept for an almost wholly neglected metaphysics.[41]

A Translations and interpretations

What follows are technical appendixes. In this first appendix, I will define some key technical notions. I begin with the notion of a translation:

---

[40] At this point, we should consider a genuinely different argument to the conclusion that all basic relations are symmetric. The argument uses DISTINCTNESS, REASONS, and these two premises in place of AUSTERITY:

FACT-ID. $R(a,b) = \breve{R}(b,a)$ for all a and b.
RELATION-ID. If both R and S are basic, and there are a and b such that $R(a,b) = S(b,a)$, then $R = S$.

Now: suppose that R is basic; then $\breve{R}$ is also basic by REASONS (arguing as before); and $R = \breve{R}$ by FACT-ID and RELATION-ID; so that $\breve{R}$ is symmetric by DISTINCTNESS. (Arguments in this ballpark are advanced by Russell (1913: 85–7), Castañeda (1975: 238–40), Armstrong (1978: 42, 94, 1997: 133–4, 143–4), Fine (2000: 2–7), and Orilia (2014: 285–6); Fine and Orilia present this as an argument against DISTINCTNESS.) Note that this argument does *not* invoke AUSTERITY; so it may not matter, to an advocate of this argument, that treating both E_{graph} and E_{alt} as basic would violate AUSTERITY.

For what it is worth, I am unmoved by this argument. The problem lies with RELATION-ID. Often, RELATION-ID is given a motivation along these lines: *suppose that both R and S are basic, and that $R(a,b) = S(b,a)$; then $R(a,b)$ has the same constituents as $S(a,b)$; so in particular $R = S$.* (Cf. Russell (1913: 85–7), Castañeda (1975: 238–40), Fine (2000: 4–5), MacBride (2014: 4), and Orilia (2014: 285).) That motivation is uncompelling. In mentioning "constituents", we are implicitly instructed to regard facts (or propositions) quasi-mereologically. At best, such a quasi-mereological approach is *optional*. (This is Trueman's (2021: 147) response to Fine; see also Leo (2013: 357–9) and Liebesman (2014: 412–3).) At worst, quasi-mereological approaches to propositions are *inconsistent*, since they court the Russell–Myhill Paradox.

[41] Special thanks to Allen Hazen and Albert Visser for many incredibly helpful suggestions, several of which are specifically referenced in other footnotes. Thanks also to Nilanjan Das, Ali Enayat, Joel David Hamkins, Johannes Korbmacher, Fraser MacBride, Rob Trueman, and José Zalarbado.

Definition A.1: Let $\mathscr{K}$ and $\mathscr{L}$ be two relational signatures. An identity-preserving translation, $* : \mathscr{K} \longrightarrow \mathscr{L}$, comprises the following pieces of information:

(1) an $\mathscr{L}$-formula $\delta_*(\mathrm{x})$, which is called the "domain formula"; and
(2) an $\mathscr{L}$-formula $\mathrm{R}^*(\mathrm{x}_1, \ldots, \mathrm{x}_n)$, for each n-place predicate $\mathrm{R} \in \mathscr{K}$.

We say that $*$ is *direct* iff $\delta_*(\mathrm{x}) :\equiv (\mathrm{x} = \mathrm{x})$. So a direct translation, in effect, only acts on the atomic predicates in $\mathscr{K}$.

We write ϕ^* for the $\mathscr{L}$-formula which results by $*$-translating the $\mathscr{K}$-formula ϕ. This is obtained recursively: condition (2) tells us how to $*$-translate atomic $\mathscr{K}$-formulas (since $*$ is identity-preserving, we do nothing to identities); then $*$ commutes with sentential connectives, and restricts quantifiers to δ_*. This recursive definition is best illustrated with an example. So, let $*$ be an identity-preserving translation from the signature of set theory; then $(\forall a \exists b \forall x (x \in b \leftrightarrow x = a))^*$ is:

$$\forall a(\delta_*(a) \rightarrow \exists b(\delta_*(a) \wedge \forall x(\delta_*(x) \rightarrow ((x \in b)^* \leftrightarrow x = a))))$$

In what follows, I often write e.g. $(\mathrm{x} \in \mathrm{y})^*$ as $\mathrm{x} \in^* \mathrm{y}$, for readability.

Note that Definition A.1 only covers translations between *relational* signatures. This is no real restriction; we can always start by replacing any n-place function-symbol with an $(n+1)$-place predicate. Definition A.1 also only covers *identity-preserving* translations. This *is* a genuine restriction; but the restriction does not matter much for my purposes, since (almost)[42] all of the translations I consider in this paper are identity-preserving.

In §4, I mentioned that translations can be applied "within" structures. Here is the point, spelled out formally:

Definition A.2: Let $\mathcal{M}$ be an $\mathscr{L}$-structure, and let $* : \mathscr{K} \longrightarrow \mathscr{L}$ be a translation. Then $*\mathcal{M}$ is the $\mathscr{K}$-structure whose domain is $\delta_*^{\mathcal{M}}$, and where $\mathrm{R}^{(*\mathcal{M})} = (\mathrm{R}^*)^{\mathcal{M}} = \{\langle a_1, \ldots, a_n \rangle : \mathcal{M} \vDash \mathrm{R}^*(a_1, \ldots, a_n)\}$, for predicate $\mathrm{R} \in \mathscr{K}$.

Having considered translations, we now consider *interpretations*. These are essentially translations that preserve theoremhood. In detail:

Definition A.3: An *interpretation* $* : \mathbf{T} \longrightarrow \mathbf{S}$ is a theorem-preserving translation of $\mathbf{T}$'s signature into $\mathbf{S}$'s signature; i.e. if $\mathbf{T} \vdash \phi$ then $\mathbf{S} \vdash \phi^*$, for any $\mathbf{T}$-sentence ϕ. An interpretation is *faithful* iff it also preserves non-theorems, i.e. $\mathbf{T} \vdash \phi$ iff $\mathbf{S} \vdash \phi^*$, for any $\mathbf{T}$-sentence ϕ. An interpretation is *direct* iff the underlying translation is direct.

Note that, if $* : \mathbf{T} \longrightarrow \mathbf{S}$ is an interpretation and $\mathcal{M} \vDash \mathbf{S}$, then $*\mathcal{M} \vDash \mathbf{T}$. As a further exercise in applying these notions, consider this result (recalling Definition 1):

[42] In fact, Proposition 2 and Lemma A.4 hold for arbitrary interpretations. For the fully general notion of an interpretation, see e.g. Visser (2008: 301) and Friedman and Visser (2014: §2.2).

Lemma A.4: Let **T** be a finitely axiomatized, essentially undecidable theory. No consistent monadic theory interprets **T**.[43]

Proof. Let $* : \mathbf{T} \longrightarrow \mathbf{O}$ be an interpretation, with **O** monadic. Define two sets of sentences, $\mathbf{T}^* := \{\phi^* : \phi \in \mathbf{T}\}$ and $\mathbf{U} := \{\phi : \mathbf{T}^* \vdash \phi^*\}$. Clearly $\mathbf{T}^* \subseteq \mathbf{O}$ and $\mathbf{T} \subseteq \mathbf{U}$. Moreover, for any formula ϕ in **T**'s signature:

$$\mathbf{U} \vdash \phi \text{ iff } \mathbf{T}^* \vdash \phi^*$$

Left-to-right holds as $*$ is a translation; right-to-left holds by definition of **U**. Since $\mathbf{T}^*$ is finite and monadic, there is a decision procedure for the right-hand-side;[44] so **U** is decidable. Since **T** is essential undecidability, **U** is inconsistent. So $\mathbf{T}^*$ is also inconsistent, and so is **O**. □

Since Q is finitely axiomatized and essentially undecidable,[45] Lemma A.4 yields Proposition 2 of §1. But Lemma A.4 also applies to many other (weak) theories, e.g.: no consistent monadic theory interprets $\mathsf{AS_e}$ (whose axioms are given in §6).

The last notions of interpretation I need are *synonymy* and (identity-preserving) *bi-interpretability*. As I explained in §4, we can gloss synonymy as: composing interpretations you get back *exactly* where you began. The rough gloss of bi-interpretability is: composing interpretations gets you back where you began *up to definable isomorphism*. Here are formal definitions:

Definition A.5: Theories **T** and **S** are *synonymous* iff there are interpretations $I : \mathbf{T} \longrightarrow \mathbf{S}$ and $J : \mathbf{S} \longrightarrow \mathbf{T}$ such that $\mathbf{T} \vdash \phi \leftrightarrow \phi^{IJ}$ and $\mathbf{S} \vdash \phi \leftrightarrow \phi^{JI}$ for all (respectively) **T**- and **S**-formulas ϕ.[46]

An identity-preserving translation $* : \mathbf{T} \longrightarrow \mathbf{T}$ is a *self-embedding* iff there is some one-place **T**-term, $\tau(\mathrm{x})$, such that all of these hold:

(1) $\mathbf{T} \vdash \forall x\, \delta_*(\tau(x))$
(2) $\mathbf{T} \vdash \forall y(\delta_*(y) \rightarrow \exists! x\, \tau(x) = y)$
(3) $\mathbf{T} \vdash \mathrm{R}(\mathrm{x}_1, \ldots, \mathrm{x}_n) \leftrightarrow \mathrm{R}^*(\tau(\mathrm{x}_1), \ldots, \tau(\mathrm{x}_n))$, for each **T**-primitive R

Theories **T** and **S** are identity-preservingly *bi-interpretable* iff there are identity-preserving interpretations $I : \mathbf{T} \longrightarrow \mathbf{S}$ and $J : \mathbf{S} \longrightarrow \mathbf{T}$ such that both IJ and JI are self-embeddings.

Clearly, synonymy entails bi-interpretability. The converse is not generally true, but this next result gives us a very useful sufficient condition.[47]

[43] This proof assumes that $*$ is identity-preserving. However, the result itself holds for arbitrary (indeed, multi-dimensional) interpretations; to generalize the proof, we simply need to augment **T** with axioms to govern identity. Thanks to Albert Visser for discussion of this.

[44] The decidability of monadic first-order logic is a textbook result; see e.g. Boolos et al. (2007: Theorem 21.6).

[45] Again, this is a textbook result; see e.g. Boolos et al. (2007: Theorem 17.5).

[46] Given translations $I : \mathscr{L}_1 \longrightarrow \mathscr{L}_2$ and $J : \mathscr{L}_2 \longrightarrow \mathscr{L}_3$, their composition is $IJ : \mathscr{L}_1 \longrightarrow \mathscr{L}_3$.

[47] This is Friedman and Visser (2014: Corollary 5.5).

Friedman–Visser Theorem: If **T** and **S** are identity-preservingly bi-interpretable, and either theory is *conceptual*, then **T** and **S** are synonymous.

The definition of a *conceptual* theroy is a little lengthy, so I will simply point the reader to Friedman and Visser (2014).

B Proving Proposition 8

I will now prove Proposition 8, which provides us with a selection of unsymmetrizable theories (see §§4–5). Specifically, I will prove clauses (2)–(4) of Proposition 8; the reader can confirm clause (1) by modifying the proof for clause (3).

Lemma B.1: DLO, the first-order theory of $(\mathbb{Q}, <)$, is unsymmetrizable.

Proof. For concision, let $Q = (\mathbb{Q}, <)$. For reductio, let **S** be symmetric and synonymous with the theory of Q. Let $\sharp$ and $\flat$ witness the synonymy. Note that $\flat Q \vDash \mathbf{S}$. Observe two facts:

(1) For any rationals $p < q$ and $r < s$, there is an automorphism on Q which maps $p \mapsto r$ and $q \mapsto s$.
(2) Every automorphism on Q is an automorphism on $\flat Q$, and vice versa.

Fact (1) is an elementary fact about Q; fact (2) follows from elementary considerations about synonymy.

Let R be any (two-place) **S**-primitive. Suppose there are distinct rationals, $p \neq q$ such that $\flat Q \vDash \mathrm{R}(p, q)$. Fix any rationals $r \neq s$. Recalling that $<$ is a total order on Q, we now show that $\flat Q \vDash \mathrm{R}(r, s)$ by considering four cases.

When $p < q$ and $r < s$: by (1), some automorphism on Q sends $p \mapsto r$ and $q \mapsto s$; by (2), the same automorphism in $\flat Q$ yields that $\flat Q \vDash \mathrm{R}(r, s)$.

When $q < p$ and $s < r$: similarly $\flat Q \vDash \mathrm{R}(r, s)$.

When $q < p$ and $r < s$: since **S** is symmetric, also $\flat Q \vDash \mathrm{R}(q, p)$. Now $\flat Q \vDash \mathrm{R}(r, s)$ as in the first case.

When $p < q$ and $s < r$: similar juggling yields $\flat Q \vDash \mathrm{R}(r, s)$.

Having covered all cases, we can generalize: for any (two-place) **S**-primitive, R, either every pair of distinct elements satisfies R in $\flat Q$, or none do. Consequently, the map f given by $f(p) = -p$ is an automorphism on $\flat Q$. So f is also an automorphism on Q, by (2). But that is absurd; f reverses Q's order. □

Lemma B.2: The first-order theory of $(\mathbb{Z}, <)$ is unsymmetrizable.

Proof. Let $\mathcal{Z} = (\mathbb{Z}, <)$. For each $k \in \mathbb{Z}$, the map $i \mapsto (i + k)$ is an automorphism on $\mathcal{Z}$. Let **S** be symmetric and synonymous with the theory of $\mathcal{Z}$. Reasoning as in Lemma B.1: for any (two-place) **S**-primitive, R, there is a set $P \subseteq \mathbb{N}$ such that, for any i and j, we have: $\flat\mathcal{Z} \vDash \mathrm{R}(i, j)$ iff $|i - j| \in P$. Now the map $f(i) = -i$ is an automorphism on $\flat\mathcal{Z}$, as $|i - j| = |f(i) - f(j)|$; so f is also, absurdly, an automorphism on $\mathcal{Z}$. □

Lemma B.3: Robinson Arithmetic, Q, is unsymmetrizable.

Proof. We start with a model, $\mathcal{M}$, of Q, described by Visser.[48] $\mathcal{M}$ comprises the natural numbers, followed by a copy of the integers, i.e. elements i_* for every $i \in \mathbb{Z}$. For every $k \in \mathbb{Z}$, the map f_k is an automorphism on $\mathcal{M}$, where:

$$f_k(i_*) = (i + k)_*, \text{ for all } i \in \mathbb{Z}$$
$$f_k(n) = n, \text{ for all } n \in \mathbb{N}$$

Let **S** be symmetric and synonymous with Q. As in Lemma B.1: for any (two-place) **S**-primitive, R, there is some $P \subseteq \mathbb{N}$ such that, for all $i, j \in \mathbb{Z}$ and all $n \in \mathbb{N}$:

- $\flat\mathcal{M} \vDash \mathrm{R}(i_*, j_*)$ iff $|i - j| \in P$; and
- $\flat\mathcal{M} \vDash \mathrm{R}(i_*, n)$ iff $\flat\mathcal{M} \vDash \mathrm{R}(j_*, n)$ iff $\flat\mathcal{M} \vDash \mathrm{R}(n, i_*)$.

Now the map given by $f(i_*) = (-i)_*$ and $f(n) = n$, for all $i \in \mathbb{Z}$ and all $n \in \mathbb{N}$, is an automorphism on $\flat\mathcal{M}$. So f is also, absurdly, an automorphism on $\mathcal{M}$. □

After proving these results, I was pleased to discover that, several decades ago, Svenonius had offered exactly the argument given in my Lemma B.2.[49] Svenonius also provided a combinatorial argument which gives another source of unsymmetrizable theories.[50] For any $n > 1$, let Φ_n be the theory of an arbitrary n-place relation. (So Φ_n has no axioms; it just amounts to specifying a signature.) We can show: *Any theory in a relational signature which directly and faithfully interprets Φ_n must have a primitive with at least n-places.* Hence Φ_n is unsymmetrizable whenever $n > 2$.

C The Graphability Theorem

In this appendix, I will state and prove the main result of this paper, the Graphability Theorem. As a corollary, this yields Proposition 10, which I discussed in §§6–8.

C.1 Stating the Graphability Theorem

I must start by stating the Graphability Theorem. This is a "proof-generated" strengthening of Proposition 10, in the sense that the strengthening emerges by scrutinizing the assumptions used in my proof-strategy. Here is the proof-strategy:[51]

[48] Visser (2008: 304–5). Constants are interpreted as follows. *Zero*: $0^{\mathcal{M}} = 0$. *Successor*: $s^{\mathcal{M}}(a) = s(a)$ if $a \in \mathbb{N}$; otherwise $s^{\mathcal{M}}(a) = a$. *Plus*: $a +^{\mathcal{M}} b = a + b$ if both $a, b \in \mathbb{N}$; otherwise $a +^{\mathcal{M}} b = \max(a, b)$. *Times*: $a \times^{\mathcal{M}} b = a \times b$ if both $a, b \in \mathbb{N}$; $a \times^{\mathcal{M}} b = 0$ if either a or b is 0; otherwise $a \times^{\mathcal{M}} b = \max(a, b)$.

[49] Svenonius (1955: Theorem 24), although he does not state the matter in terms of synonymy (the notion had not yet been invented). Thanks to Allen Hazen for alerting me to Svenonius's paper.

[50] Svenonius (1955: Theorems 22–3).

[51] This strategy builds on the work of many people. Hodges (1993: Theorem 5.5.1) shows that any theory in a finite signature is (multi-dimensionally) bi-interpretable with some graph theory. (Hodges credits Lavrov 1963 with the result and the argument.) Using a theory which allows for some coding, as per (G2), I can me to turn Hodges' multi-dimensional interpretation into a one-dimensional interpretation by treating tuples-as-sets (and slightly modifying Hodges' coding to

- Working within some theory, **T**, I describe a graph-theoretic universe which encodes all of **T** (see §C.3).
- I then define a graph theory, $\mathbf{T}_{\mathsf{graph}}$, which axiomatizes this graph-universe (see §C.4).
- Finally, I show that **T** and $\mathbf{T}_{\mathsf{graph}}$ are bi-interpretable (see §§C.5–C.6); their synonymy will then follow from the Friedman–Visser Theorem (see the end of §A). So $\mathbf{T}_{\mathsf{graph}}$ will witness **T**'s graphability.

Clearly, my proof-strategy requires both that the Friedman–Visser Theorem applies to **T**, and that **T** is rich enough to carry out my proposed coding. More explicitly, here are my required assumptions:

(G1) **T** is a first-order theory with finitely many non-logical primitives;
(G2) **T** has a universal, canonical, non-surjective, ordered-pairing operation; and
(G3) **T** is *conceptual* (in the Friedman–Visser sense).

Condition (G3) is just the condition required for the Friedman–Visser Theorem. Conditions (G1)–(G2) allow me to execute a certain amount of coding. Condition (G1) speaks for itself, but (G2) is quite compressed; here it is, spelled out fully. There is a closed **T**-term, 0, and a **T**-term with two free variables, $\langle \mathrm{x}, \mathrm{y} \rangle$, such that:

$$\mathbf{T} \vdash \forall a \forall b \forall a' \forall b' \big(\langle a, b \rangle = \langle a', b' \rangle \rightarrow (a = a' \wedge b = b') \big)$$
$$\mathbf{T} \vdash \forall a \forall b \; 0 \neq \langle a, b \rangle$$

Spelling this out: $\langle \mathrm{x}, \mathrm{y} \rangle$ is our *ordered-pairing operation*; that operation is *universal*, in that $\langle a, b \rangle$ exists for any a and b; it is *canonical*, in that, given a and b (in that order), we can uniquely pick out some (canonical) ordered-pair as $\langle a, b \rangle$; and it is canonically *non-surjective*, in that we can uniquely pick out 0 as a (canonical) non-ordered-pair.

I am now in a position to state the main result of this paper:

Graphability Theorem: If **T** meets conditions (G1)–(G3), then **T** is graphable.

I stated Proposition 10 in the main text, rather than the Graphability Theorem, because directly interpreting $\mathsf{AS_e}$ is a little more "concrete" than conditions (G2)–(G3). However, the Graphability Theorem is the strongest result possible which can be obtained by my proof strategy.

The Graphability Theorem immediately entails Proposition 10. In particular, if **T** directly interprets $\mathsf{AS_e}$, then **T** is certainly conceptual,[52] and **T** has a universal, canonical, non-surjective, ordered-pairing operation, in the form of its (direct) interpretation of the Kuratowski-definition of ordered-pairs. The Graphability Theorem

ensure it is identity-preserving). Using a theory conceptual theory allows me to apply the Friedman–Visser Theorem.

Many thanks to: Joel David Hamkins, who gave me an excellent proof-strategy which got me started on this; Ali Enayat, who drew my attention to Hodges' work after I had (clunkily) proved an earlier version of Proposition 10; and Albert Visser, who gave me all sorts of helpful advice.

[52] Since **T** directly interprets $\mathsf{AS_e}$, it directly interprets the weaker theory AS, so that **T** is *sequential* (by definition); and every sequential theory is conceptual (see Friedman and Visser 2014: §3).

strictly improves on Proposition 10, though, because there are theories which meet conditions (G1)–(G3) but which do not directly interpret $\mathsf{AS_e}$.[53]

C.2 Tuples and numerals

For the result of this appendix, I will use **T** for an arbitrary theory which is assumed to meet conditions (G1)–(G3). Without loss of generality, I will make a few further assumptions:

- All of **T**'s axioms are closed sentences.[54]
- All of **T**'s non-logical primitives are predicates.[55] I will let $\mathrm{R}_1, \ldots, \mathrm{R}_N$ enumerate them; for each $1 \leq i \leq N$, the predicate R_i has σ_i-places.
- Some **T**-primitive delivers the ordered-pairing operation of condition (G2).[56]

When working in **T**, I will write $\langle a, b\rangle$ for the (canonical) ordered-pair whose first entry is a and whose second is b; canonical triples will be canonical pairs whose second element is a canonical pair, i.e. $\langle a, b, c\rangle = \langle a, \langle b, c\rangle\rangle$; etc.

I can also provide a canonical notion of a **T**-*numeral*. By condition (G2), there is some canonical non-pair; let this be our 0. For each n, we then specify that $n + 1$ will be $\langle n, 0\rangle$. Using these definitions, we can now establish an important result. (The parenthetical note, after the announcement of a result, indicate the theory for which the result holds.)

Lemma C.1 (**T**; schematic for n)**:** 0, 1, 2, 3, . . . , n all exist, and are all distinct.

Proof. An easy (metatheoretic) induction, using condition (G2). □

C.3 Coding a graph-universe

We have seen that **T** can implement arbitrary (finite) tuples and arbitrary numerals. Using these coding tools repeatedly, and without further comment, I will now outline a strategy for describing a graph-universe within **T** which encodes **T** itself.

Coding the domain. For each object, x, we must identify some (unique) object, $\Gamma(x)$, which will go proxy for x as a node in our graph. I stipulate the following:

$$\Gamma(x) :\equiv \langle x, 0, 0\rangle$$

[53] Albert Visser (private communication) suggests the following example. Let $\mathsf{AS_p}$, for Adjunctive Set theory with (canonical) pairs, have these three axioms:

$$\forall x\, x \notin \emptyset$$
$$\forall a \forall b \exists c \forall x (x \in c \leftrightarrow (x \in a \vee x = b))$$
$$\forall a \forall b \forall x (x \in a{+}b \leftrightarrow (x = a \vee x = b))$$

So, $\mathsf{AS_p}$ has three non-logical primitives: a constant, $\emptyset$; a two-place relation, $\in$; and a two-place function-symbol, $+$. It is easy to confirm that $\mathsf{AS_p}$ meets conditions (G1)–(G3). Moreover, $\mathsf{AS_e}$ directly interprets $\mathsf{AS_p}$, but not conversely.

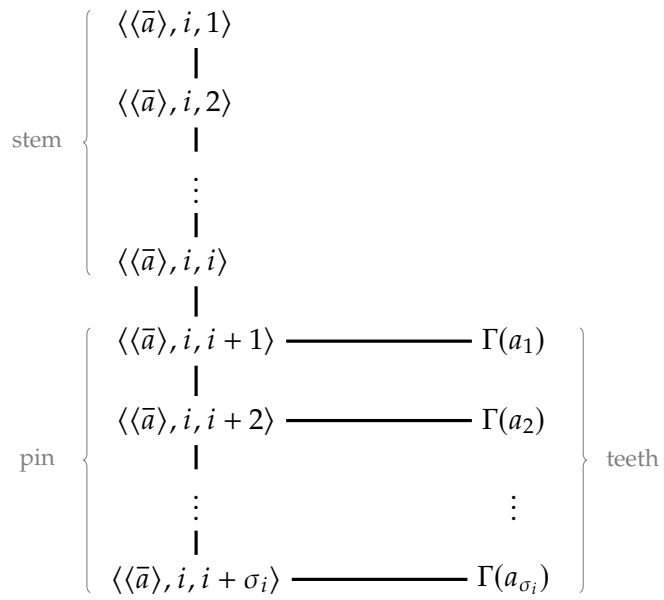
[54] No generality is lost, since we can just take their universal closures.

[55] No generality is lost, via usual algorithms for replacing function-symbols with predicates.

[56] Specifically, for some $1 \leq l \leq N$, **T** proves both $\forall a \forall b \exists! c \mathrm{R}_l(a, b, c)$ and $(\mathrm{R}_l(a, b, c) \wedge \mathrm{R}_l(a', b', c)) \rightarrow (a = a' \wedge b = b')$. No generality is lost, since we are ultimately only interested in synonymy.

Note that Γ is functional, by condition (G2). Indeed, Γ is also obviously *injective*; that is, given $\Gamma(x)$, we can uniquely recover x.

Coding each R_i-fact. I code each fact of the form $\mathrm{R}_i(a_1, \ldots, a_{\sigma_i})$ with a unique graph-theoretic configuration, which I call a *key*. In brief: given a *key*, we inspect the length of its stem to determine which R_i it encodes; we then determine which objects are related by R_i, and in what order, by running along the *teeth* on the key's *pin*. Here is a picture of the key which would encode that $\mathrm{R}_i(a_1, \ldots, a_{\sigma_i})$ (I abbreviate sequences with overlining, so $\langle \overline{a} \rangle$ is just $\langle a_1, \ldots, a_{\sigma_i} \rangle$):



Let me now explain how the various parts of this key act together, to encode the fact that $\mathrm{R}_i(a_1, \ldots, a_{\sigma_i})$.

The *stem* of a key which encodes an R_i-fact is a path of length i. The stem thereby tells us which R_i we are dealing with. The stem then connects to the *pin*, which is a path of length $\sigma_i - 1$. We can treat each node in the pin as an "argument-slot" for R_i. The kth node in the pin has an edge to the kth *tooth*. The kth *tooth* itself is just the proxy for a_k, i.e. it is just $\Gamma(a_k)$. So the key's teeth tell us which elements occupy which of R_i's argument slots.

It should now be clear that, given any R_i-facts, we can uniquely describe the keys which encode those facts; and given any keys, we can decode them to discover which R_i-facts they encode. Of course, this all remains at the intuitive, hand-waving, level; the next step is to formalize it properly.

Just before I do that, though, I want to pause briefly, to emphasise that this construction involves plenty of choices which are essentially arbitrary. For example: we would obtain a strictly different construction, if we enumerated **T**'s predicates differently (this would affect the length of the "stem" associated with each predicate). More deeply: there is no need to use "keys" to code up each R_i-fact; plenty of other "shapes" would do. In short: I have chosen a coding-strategy which is convenient for the proof of the Graphability Theorem; I do not claim that this coding-strategy reflects any deep metaphysics (see §8).

I now return to the aim of formalizing my coding-strategy. This requires laying down various definitions within **T**. (For readability, I will write these in a semi-formal way; but all the definitions are obviously fully first-orderizable.) I first lay down the "domain" for our graph-like objects. Recall that **T**-predicates are enumerated $\mathrm{R}_1, \ldots, R_N$. So I define:

$$\begin{aligned} \mathit{Node}(\mathrm{x}) :\equiv\ & \mathrm{x} \text{ is some triple } \langle a, i, j\rangle \text{, and either:} \\ & 0 = i = j \text{; or} \\ & 1 \leq i \leq N \text{ and } 1 \leq j \leq i + \sigma_i \text{ and } \exists u_1 \ldots \exists u_{\sigma_i}(a = \langle \overline{u}\rangle \wedge \mathrm{R}_i(\overline{u})) \end{aligned}$$

The case when $\mathrm{x} = \langle a, 0, 0\rangle$ covers the case of $\Gamma(\mathrm{x})$; these, recall, are our teeth. The next line of this definition covers our stems and pins. I now define an "edge" relation, to link the stems, pins and teeth:

$$\begin{aligned} C(\mathrm{x}, \mathrm{y}) :\equiv\ & \mathit{Node}(\mathrm{x}) \text{ and } \mathit{Node}(\mathrm{y}) \text{ and where } \mathrm{x} = \langle a_{\mathrm{x}}, i_{\mathrm{x}}, j_{\mathrm{x}}\rangle \text{ and } \mathrm{y} = \langle a_{\mathrm{y}}, i_{\mathrm{y}}, j_{\mathrm{y}}\rangle\text{:} \\ & 1 \leq i_{\mathrm{x}} = i_{\mathrm{y}} \text{ and } a_{\mathrm{x}} = a_{\mathrm{y}} \text{ and } |j_{\mathrm{x}} - j_{\mathrm{y}}| = 1 \text{; or} \\ & 1 \leq i_{\mathrm{x}} \text{ and } 0 = i_{\mathrm{y}} \text{ and } a_{\mathrm{y}} \text{ is } a_{\mathrm{x}}\text{'s } (j_{\mathrm{x}} - i_{\mathrm{x}})^{\text{th}}\text{-entry; or} \\ & 1 \leq i_{\mathrm{y}} \text{ and } 0 = i_{\mathrm{x}} \text{ and } a_{\mathrm{x}} \text{ is } a_{\mathrm{y}}\text{'s } (j_{\mathrm{y}} - i_{\mathrm{y}})^{\text{th}}\text{-entry.} \end{aligned}$$

The first clause of this definition covers the case when x and y are adjacent parts of the stem/pin in a single key. The second clause covers the case when x is part of a stem which should be linked to y as a tooth, e.g. $\mathrm{x} = \langle\langle b_1, b_2, b_3\rangle, i, i+2\rangle$ and $\mathrm{y} = \langle b_2, 0, 0\rangle$. The third line reverses the roles of x and y to ensure that C is symmetric.

This completes the coding. We have now defined a graph-universe within **T** which encodes **T**.

C.4 The formal theory $\boldsymbol{T}_{\mathit{graph}}$

The next step is provide an axiomatic graph theory which governs this graph-universe, $\mathbf{T}_{\mathsf{graph}}$. This is possible, because we can determine a node's *kind* (in the graph-universe) just by considering its degree and the degree of nodes to which it is connected.[57] That is what the following result tells us:

Lemma C.2 (T): Let $x = \langle a, i, j\rangle$ be a *Node*-entity.

(1) $i = 0$ iff $\deg_C(x) > 3$.
(2) if $i > 0$:
  (i) if $j = 1$, then $\deg_C(x) = 1$;
  (ii) if $i < j \leq i$ or $j = i + \sigma_i$, then $\deg_C(x) = 2$;
  (iii) if $i < j < i + \sigma_i$, then $\deg_C(x) = 3$.

[57] As usual, the degree of a node x is the cardinality of nodes to which x has edges. I write this as $\deg_C(x)$, when C is the edge relation. Since all the numbers involved here are small and finite, all this remains first-orderizable.

Proof. Inspecting the definitions yields (2). This also gives right-to-left of (1). For left-to-right of (1), suppose $i = 0$. Let $\mathbf{T}$'s primitive pairing-predicate be R_l (see the assumptions of §C.2). Using Lemma C.1, let $c_k = \langle a, k\rangle$ for each $0 \leq k \leq 3$; these are all distinct, as are all of $\langle\langle a, k, c_k\rangle, l, l+1\rangle$, and $\Gamma(a) = x$ has an edge to each just by considering each key which encodes $\mathrm{R}_l(a, k, c_k)$. □

Case (1) covers all possible teeth. Then within (2): case (i) is the first node of a stem; case (ii) covers the remaining nodes of a stem and the last node of a pin; and case (iii) covers the remaining nodes of a pin. Note that the stem-nodes in (ii) will be discernible from the pin-node, since only the pin-node will have an edge to a tooth, i.e. to a node with degree > 3.

Using this insight, we can write define formulas which specify our different kinds of node.

$$\begin{aligned}
\mathit{Tooth}(\mathrm{x}) &:\equiv \deg_E(\mathrm{x}) > 3 \\
\mathit{Stem}(\mathrm{x}) &:\equiv 1 \leq \deg_E(\mathrm{x}) \leq 2 \text{ and } \neg\exists v(\mathit{Tooth}(v) \wedge E(\mathrm{x}, v)) \\
\mathit{Pin}(\mathrm{x}) &:\equiv 2 \leq \deg_E(\mathrm{x}) \leq 3 \text{ and } \exists v(\mathit{Tooth}(v) \wedge E(\mathrm{x}, v))
\end{aligned}$$

Building up a little complexity, we can define the first node in the stem of a key which codes some R_i-fact (for any $1 \leq i \leq N$):

$$\begin{aligned}
\mathit{Key}_i^1(\mathrm{x}) :\equiv\ & \mathit{Stem}(\mathrm{x}) \text{ and } \deg_E(\mathrm{x}) = 1 \text{ and} \\
& \text{there is a path, starting with x, followed by } i-1 \text{ } \mathit{Stem}\text{s,} \\
& \text{then followed by } \sigma_i \text{ } \mathit{Pin}\text{s, the last of which has degree 2}
\end{aligned}$$

And then we can easily individuate the other parts of such a key (for any $1 \leq i \leq N$ and $1 < j \leq i + \sigma_i$):

$$\begin{aligned}
\mathit{Key}_i^j(\mathrm{x}) :\equiv\ & (\mathit{Stem}(\mathrm{x}) \vee \mathit{Pin}(\mathrm{x})) \text{ and there is a path of length } j-1, \\
& \text{comprising only } \mathit{Stem}\text{s and } \mathit{Pin}\text{s, from some } \mathit{Key}_i^1 \text{ to x}
\end{aligned}$$

Finally, we can use this to define a graph-theoretic expression which will act as a way to code R_i-facts, for each $1 \leq i \leq N$. Roughly, we need σ_i-many teeth, connected appropriately to key-components. More precisely:

$$\begin{aligned}
\mathrm{R}_i^J(\mathrm{x}_1, \ldots, \mathrm{x}_{\sigma_i}) :\equiv\ & \textstyle\bigwedge_{j=1}^{\sigma_i} \mathit{Tooth}(\mathrm{x}_j) \wedge \exists v_1 \ldots v_{i+\sigma_i} \big(\bigwedge_{j=1}^{i+\sigma_i} \mathit{Key}_i^j(v_j) \wedge \\
& \textstyle\bigwedge_{j=1}^{i+\sigma_i-1} E(v_j, v_{j+1}) \wedge \bigwedge_{j=1}^{\sigma_i} E(v_{i+j}, \mathrm{x}_j)\big)
\end{aligned}$$

The superscript "J" indicates that we will use these formulas to define a direct translation, J, from $\mathbf{T}$ (see Definition A.1). The only further component we require is a domain-formula, δ_J; this is simply *Tooth*. I can now use J to define $\mathbf{T}_{\mathsf{graph}}$ as the theory with these axioms:

- ϕ^J, for any $\mathbf{T}$-axiom ϕ
- E is symmetric and irreflexive

- everything is either a *Tooth* or some Key_i^j with $1 \leq i \leq N$ and $i+1 \leq j \leq \sigma_i$
- there are no edges between *Tooth*s
- "keys are unique", i.e., given keys comprising $u_1, \ldots, u_{i+\sigma_i}$ and $v_1, \ldots, v_{i+\sigma_i}$, and where both keys witness that $\mathrm{R}_i^J(a_1, \ldots, a_{\sigma_i})$, we have $\bigwedge_{j=1}^{i+\sigma_i} u_j = v_j$.

This completes the definition of $\mathbf{T}_{\mathsf{graph}}$, and formalizes our graph-theoretic universe.

C.5 *Bi-interpreting $\mathbf{T}_{graph}$ in $\mathbf{T}$*

The first clause of $\mathbf{T}_{\mathsf{graph}}$'s axioms immediately gives us the following:

Lemma C.3: $J : \mathbf{T} \longrightarrow \mathbf{T}_{\mathsf{graph}}$ is an interpretation.

The other axioms of $\mathbf{T}_{\mathsf{graph}}$ ensure that there are no *other* edges or nodes than we specified in our $\mathbf{T}$-defined graph-universe. Indeed, they ensure that $\mathbf{T}$ and $\mathbf{T}_{\mathsf{graph}}$ are bi-interpretable. But, to show this, I must define an identity-preserving translation, I, as an "inverse" to J; i.e. such that JI is a self-embedding (see Definition A.5).

The definition of I is unsurprising. The domain-formula, δ_I, is just *Node*, and $\mathbf{T}_{\mathsf{graph}}$'s single predicate is translated thus:

$$E^I(\mathrm{x}, \mathrm{y}) :\equiv C(\mathrm{x}, \mathrm{y})$$

We now have some quick results showing that this behaves as we would like. Our first result, in effect, says that our keys do exactly the coding job we demanded that we built them to do:

Lemma C.4 ($\mathbf{T}$): $\mathrm{R}(a_1, \ldots, a_{\sigma_i})$ iff $\mathrm{R}^{JI}(\Gamma(a_1), \ldots, \Gamma(a_{\sigma_i}))$

Proof. The definitions were constructed precisely to ensure this. □

Lemma C.5 ($\mathbf{T}$): Γ witnesses that JI is a self-embedding.

Proof. Γ exhausts $Tooth^I$ by Lemma C.2. Γ is injective. And Γ preserves structure by Lemma C.4. □

Lemma C.6: $I : \mathbf{T}_{\mathsf{graph}} \longrightarrow \mathbf{T}$ is an interpretation.

Proof. By Lemma C.5, $\mathbf{T}$ proves ϕ^{JI} for any $\mathbf{T}$-axiom ϕ. Then $\mathbf{T}$ proves the I-translations of $\mathbf{T}_{\mathsf{graph}}$'s other axioms, using Lemma C.2. □

C.6 *Bi-interpreting $\mathbf{T}$ in $\mathbf{T}_{graph}$, and synonymy*

We already know that J is an interpretation; so it just remains to show that it forms the other half of our bi-interpretation, i.e. that IJ is a self-embedding. To this end, I need some abbreviations.

Recall from §C.2 that $\mathbf{T}$ provides us with (canonical) numerals and ordered-tuples. Using our interpretation $J : \mathbf{T} \longrightarrow \mathbf{T}_{\mathsf{graph}}$, we can consider the J-translation

of the (canonical) definitions of numerals and ordered-tuples. I write "$\langle\!\langle x, i^J, j^J \rangle\!\rangle$" for the translation of the definition of the (canonical) triple whose first element is x and whose second and third are the appropriate (canonical) numerals.

Next, note that if $Key_i^j(u)$, then since "keys are unique", we can explicitly define (in $\mathbf{T}_{\text{graph}}$) those *Tooth*s, $a_1, \ldots, a_{\sigma_i}$, which are connected (in that order) to the pin of the key to which u belongs. We can then uniquely define the J-translation of the ordered-tuple whose entries are those *Tooth*s, i.e. $\langle\!\langle a_1, \ldots, a_{\sigma_i} \rangle\!\rangle$. Call this $\mathring{u}$, recognising that this term abbreviates a very lengthy $\mathbf{T}_{\text{graph}}$-definition.

Using all this notation, I define a map within $\mathbf{T}_{\text{graph}}$:

$$\Delta(u) :\equiv \begin{cases} \langle\!\langle u, 0^J, 0^J \rangle\!\rangle & \text{if } Tooth(u) \\ \langle\!\langle \mathring{u}, i^J, j^J \rangle\!\rangle & \text{if } Key_i^j(u) \end{cases}$$

Lemma C.7 ($\mathbf{T}_{\text{graph}}$): Δ witnesses that IJ is a self-embedding

Proof. Since everything is either a *Tooth* or some Key_i^j, Δ is total.

Δ *is injective.* Suppose $\Delta(x) = \Delta(y)$. Evidently, x and y are either both *Tooth*-nodes, or both Key_i^j-nodes (for the same i and j). If both are *Tooth*-nodes, injectivity is immediate; otherwise, injectivity holds as "keys are unique".

Δ *exhausts* $Node^J$. Suppose $Node^J(x)$. If $x = \langle\!\langle u, 0^J, 0^J \rangle\!\rangle$ for some *Tooth*-node u, then $\Delta(u) = x$. If $x = \langle\!\langle y, i^J, j^J \rangle\!\rangle$ with $i^J \neq 0^J$, then there are $a_1, \ldots, a_{\sigma_i}$ such that $y = \langle\!\langle \bar{a} \rangle\!\rangle$ and $\mathrm{R}_i^J(\bar{a})$; so there is some key whose teeth are $\bar{a}$; let u be the element of that key such that $Key_i^j(u)$. Now $\mathring{u} = \langle\!\langle \bar{a} \rangle\!\rangle = y$ and $\Delta(u) = x$.

It remains to show that $E(x, y)$ iff $C^J(\Delta(x), \Delta(y))$. There are three cases to consider; I silently invoke Lemma C.2^J throughout.

Case when $Tooth(x)$ *and* $Tooth(y)$. Now $\neg E(x, y)$. Moreover, $\Delta(x) = \langle\!\langle x, 0^J, 0^J \rangle\!\rangle$ and $\Delta(y) = \langle\!\langle y, 0^J, 0^J \rangle\!\rangle$, so that $\neg C^J(\Delta(x), \Delta(y))$.

Case when $Tooth(x)$ *and* $Key_i^j(y)$. Now $\Delta(x) = \langle\!\langle x, 0^J, 0^J \rangle\!\rangle$ and $\Delta(y) = \langle\!\langle \mathring{y}, i^J, j^J \rangle\!\rangle$. So: $E(x, y)$ iff x is the $(j - i)^{\text{th}}$ element of $\mathring{y}$; i.e. iff $C^J(\Delta(x), \Delta(y))$.

Case when $Key_{i_1}^{j_1}(x)$ *and* $Key_{i_2}^{j_2}(y)$. Now $\Delta(x) = \langle\!\langle \mathring{x}, i_1^J, j_1^J \rangle\!\rangle$ and $\Delta(y) = \langle\!\langle \mathring{y}, i_2^J, j_2^J \rangle\!\rangle$. So: $E(x, y)$ *iff* $\mathring{x} = \mathring{y}$ and $i_1 = i_2$ and $|j_1 - j_2| = 1$; *iff* $C^J(\Delta(x), \Delta(y))$. □

We now have all the pieces to obtain our desired synonymy:

Lemma C.8: $\mathbf{T}$ and $\mathbf{T}_{\text{graph}}$ are synonymous.

Proof. Both $J : \mathbf{T} \longrightarrow \mathbf{T}_{\text{graph}}$ and $I : \mathbf{T}_{\text{graph}} \longrightarrow \mathbf{T}$ are identity-preserving interpretations, by Lemmas C.3 and C.6. Moreover, these form a bi-interpretation, by Lemmas C.7 and C.5. Furthermore, $\mathbf{T}$ is conceptual by condition (G3). So the Friedman–Visser Theorem applies, and $\mathbf{T}$ and $\mathbf{T}_{\text{graph}}$ are synonymous. □

Since $\mathbf{T}_{\text{graph}}$ witnesses that $\mathbf{T}$ is graphable, we obtain the Graphability Theorem. As in §C.1, Proposition 10 follows immediately.

With the main result proved, let me note an amusing curiosity.

Corollary C.9: If **T** meets conditions (G1)–(G3), then the graphability of **T** can be witnessed with a theory which states "the graph has diameter 2".

Proof sketch. When describing a graph-universe in **T** which encodes **T** itself, proceed as in §C.3, but with exactly one extra node, $\star = \langle 0, N+1, 0\rangle$. Tweak the definition of *Node* to cover this triple; and tweak the definition of C so that $\star$ has an edge to *every* other node. This will ensure that the graph has diameter 2. Moreover, $\star$ is uniquely individuated as *the* node with an edge to every other node, so an extension of Lemma C.2 applies. The remainder of the proof now goes through as before, with obvious small tweaks. □

D Set theories with finite models

The Graphability Theorem does not apply to any theories which have finite models. (This is immediate from Lemma C.1.) So, given a theory with finite models, I do not have a method for showing that it is graphable. Nevertheless, for certain set theories with finite models, I do have a method which allows us to find synonymous symmetric theories.

Recall the idea of a hierarchy of sets, according to which:

> Sets are arranged in stages. Every set is found at some stage. At any stage S: for any things, each of which is either a set found before S or an urelement, we find a set whose members are exactly those things. We find nothing else at S.[58]

This story leaves open how many urelements there are (maybe there are none); and, beyond implicitly assuming that there is at least one stage, it leaves open how many stages there are.

Demonstrably, this story is fully axiomatized by the theory LTU, for Level Theory with Urelements.[59] Here is a sketch of LTU. It has two primitives, "$\in$" and "*Set*". Working in LTU, we can explicitly define a one-place predicate, "*Lev*", where we gloss "$Lev(x)$" as "x is a level". So defined, the levels provably act as proxies for the stages of the hierarchy. Moreover, we can prove in LTU that the levels are well-ordered. This allows us to define an operator, "ℓ", such that, intuitively, ℓa is the first level at which a occurs. And we can show that this all works exactly as one hope. For example, LTU proves: if $a \in b$, then $\ell a \in \ell b$.

For present purposes, the important result is that LTU is synonymous with a symmetric theory which uses exactly *two* non-logical primitives. This is Theorem D.4, below. My proof-strategy simply tweaks a result due to Hazen, concerning ZFU. Write $\ell a = \ell b$ to say that a and b have the same rank. (This is exactly how to spell out the idea within LTU, and the same definition works verbatim for ZFU.) Now, Hazen's observation is that ZFU + "there is some non-set" proves:

$$a \in b \text{ iff } (a \in b \vee b \in a) \wedge \exists u(\ell u = \ell b \wedge \forall v(\ell v = \ell a \rightarrow (u \in v \vee v \in u)))$$

[58] This particular formulation of the story is from Button (2021: §A).

[59] See Button (2021). I present both first-order and second-order versions of LTU; for reasons given in footnote 6, I use first-order LTU in this appendix.

My proof of Theorem D.4 builds on Hazen's observation in three ways: we can use LTU instead of ZFU; we can use some coding tricks to drop the assumption "there is some non-set"; and we can parlay this into a synonymy.[60]

In particular, my aim is to define a symmetric theory, $\mathsf{LTU}_{\mathsf{sym}}$, which is synonymous with LTU. $\mathsf{LTU}_{\mathsf{sym}}$ will have two primitives, "E" and "*Tog*". But my strategy is back-to-front: I will first explain how LTU is to simulate these new primitives; I will then use this simulation to reverse-engineer $\mathsf{LTU}_{\mathsf{sym}}$.

I start by defining a direct translation, $\flat$, from the signature $\{E, Tog\}$ to the signature $\{Set, \in\}$, as follows:

$$E^\flat(\mathrm{x}, \mathrm{y}) :\equiv (\mathrm{x} \in \mathrm{y} \vee \mathrm{y} \in \mathrm{x} \vee \mathrm{x} = \mathrm{y} = \emptyset)$$
$$Tog^\flat(\mathrm{x}, \mathrm{y}) :\equiv (\ell \mathrm{x} = \ell \mathrm{y})$$

So, "E" is interpreted as the symmetric closure of membership, with an additional self-loop from $\emptyset$ to itself; and "*Tog*" is interpreted as stating that x and y enter the hierarchy together.

For readability, I will now lay down some explicit definitions, which use only the new primitives, "E" and "*Tog*":

$$\begin{aligned} Sin(\mathrm{x}, \mathrm{y}) &:\equiv E(\mathrm{x}, \mathrm{y}) \wedge \mathrm{x} \neq \mathrm{y} \\ First(\mathrm{x}) &:\equiv \forall v(Tog(v, \mathrm{x}) \rightarrow v = \mathrm{x}) \\ Bef(\mathrm{x}, \mathrm{y}) &:\equiv \exists u(Tog(u, \mathrm{y}) \wedge \forall v(Tog(v, \mathrm{x}) \rightarrow Sin(u, v))) \\ \mathrm{x}\ \varepsilon\ \mathrm{y} &:\equiv (\neg E(\mathrm{y}, \mathrm{y}) \wedge (First(\mathrm{y}) \rightarrow E(\mathrm{x}, \mathrm{x})) \wedge Sin(\mathrm{x}, \mathrm{y}) \wedge Bef(\mathrm{x}, \mathrm{y})) \end{aligned}$$

Mnemonically, the idea is that: $Sin^\flat$ is the symmetric closure of membership; $First^\flat$ holds of the first two levels (when the hierarchy is pure); $Bef^\flat$ indicates that the one set enters the hierarchy before the other; and $\varepsilon^\flat$ turns out to be membership itself. The next few results vindicate these mnemonics:

Lemma D.1 (LTU)**:**

(1) $E^\flat(a, a)$ iff $a = \emptyset$; and
(2) $Sin^\flat(a, b)$ iff $a \in b \vee b \in a$; and
(3) if $\ell a \in \ell b$, then $Bef^\flat(a, b)$.

[60] Dorr (2004: 182–3) reports and discusses this result, in this context. Unlike Dorr, though, I present my result as a synonymy. This difference is dialectically significant.

Roughly put: Dorr suggests that Hazen's result might allow us to eliminate membership in favour of some symmetric relations; but he helps himself to membership in explaining what those symmetric relations are; and MacBride (2015: 192–4) complains that this is "circular". Whether or not this complaint tells against Dorr, it raises no problems for me.

I present a fully axiomatic, symmetric, theory, $\mathsf{LTU}_{\mathsf{sym}}$, which is synonymous with LTU. Admittedly, my presentation of $\mathsf{LTU}_{\mathsf{sym}}$ makes mention of LTU, but this is only for ease of comprehension. Indeed, we can entirely eliminate LTU in favour of $\mathsf{LTU}_{\mathsf{sym}}$, and the elimination can be carried out in a purely mechanical fashion. Consequently, there can be no damaging threat of circularity. (This can all be fruitfully compared with the lesson of §1: Gottfried would have nothing to fear from an accusation of circularity, if only—*per impossibile*—he could show how to theorize about relations in monadic terms.)

To repeat some morals from §8: this *only* rebuts the threat of circularity. Much more work would be needed, to show that the relations which $\mathsf{LTU}_{\mathsf{sym}}$ treats as primitive are more metaphysically basic than set-membership.

Proof. Only (3) is non-trivial. Suppose $\ell a \in \ell b$. Note that $\ell\ell b = \ell b$. Moreover if $\ell v = \ell a$ then $v \subseteq \ell a \in \ell b$, so that $v \in \ell b$. Hence ℓb witnesses that $Bef^\flat(a, b)$. □

Lemma D.2 (LTU)**:** If $\forall x\ Set(x)$, then both of these hold:

(1) $First^\flat(a)$ iff $a = \emptyset \vee a = \{\emptyset\}$; and
(2) if $Bef^\flat(a, b)$, then either: (i) $\ell a \in \ell b$; or (ii) $\ell a = \{\emptyset\}$ and $b = \emptyset$; or (iii) $\ell a = \{\emptyset, \{\emptyset\}\}$ and $b = \{\emptyset\}$.

If $\exists x \neg Set(x)$, then:

(3) $\neg First^\flat(a)$; and
(4) if $Bef^\flat(a, b)$, then $\ell a \in \ell b$.

Proof. I will assume that $\forall x\ Set(x)$, and prove (1)–(2); I leave it to the reader to prove that (3)–(4) under the assumption that $\exists x \neg Set(x)$.

Concerning (1). $First^\flat(a)$ states that exactly one object shares a's level; this can happen only in the first two levels of the pure hierarchy of sets.

Concerning (2). Let c witness that $Bef^\flat(a, b)$, i.e. $\ell c = \ell b$ and $\forall v(\ell v = \ell a \rightarrow (c \in v \vee v \in c))$. Note immediately that $c \in a \vee a \in c$. We now reason by cases.

When $\ell a = \emptyset$. Now $a = \emptyset$, so $c \notin a$ and hence $a \in c$; so $\ell a \in \ell c = \ell b$, i.e. (i).

When $\ell a = \{\emptyset\}$. If $\ell a \in \ell b$ we have (i); so suppose $\ell a \notin \ell b = \ell c$. Now $a \notin c$ and hence $c \in a$, i.e. $c = \emptyset$. Since $\ell c = \ell b$, also $b = \emptyset$, i.e. (ii).

When $\ell a = \{\emptyset, \{\emptyset\}\}$. As before: supposing $\ell a \notin \ell b$, we find that $c \in a$, i.e. either $c = \emptyset$ or $c = \{\emptyset\}$. For reductio, suppose $c = \emptyset$; where $d = \{\{\emptyset\}\}$, note that $\ell d = \ell a$, but $c \notin d \wedge d \notin c$, contradicting our choice of c. So $c = \{\emptyset\}$ after all. Since $\ell c = \ell b$, also $b = \{\emptyset\}$, i.e. (iii).

All other cases. Let $d = \ell a \setminus \{c\}$. Now $\ell d = \ell a$: for if $c \notin \ell a$, then $d = \ell a$; and if $c \in \ell a$, then ℓa is sufficiently well-populated that there is some $c' \in \ell a$ such that $\ell c \subseteq \ell c'$. By choice of c, we have $c \in d \vee d \in c$; so $d \in c$ by choice of d. So $\ell a = \ell d \in \ell c = \ell b$, i.e. (i). □

Lemma D.3 (LTU)**:** $a \in b$ iff $a\ \varepsilon^\flat\ b$.

Proof. In what follows, suppose $\forall x\ Set(x)$. The case when $\exists x \neg Set(x)$ is similar but easier and I leave it to the reader.

Left-to-right. Suppose $a \in b$. Then $\ell a \in \ell b$. So $Bef^\flat(a, b)$ by Lemma D.1. Also $\neg E^\flat(b, b)$ and $Sin^\flat(a, b)$. And $(First(b) \rightarrow E(a, a))^\flat$ by Lemma D.2.1.

Right-to-left. Suppose $a\ \varepsilon^\flat\ b$. Since $Bef^\flat(a, b)$, one of cases (i)–(iii) of Lemma D.2.2 holds. Since $\neg E^\flat(b, b)$, i.e. $b \neq \emptyset$, it is not case (ii). Since $(First(b) \rightarrow E(a, a))^\flat$, i.e. $(b = \emptyset \vee b = \{\emptyset\}) \rightarrow a = \emptyset$ by Lemma D.2.1, it is not case (iii). So (i) holds, i.e. $\ell a \in \ell b$. Hence $b \notin a$. So $a \in b$, because $Sin^\flat(a, b)$. □

Summarizing: working in LTU, we can find two symmetric relations, $E^\flat$ and $Tog^\flat$, and we can use them to redefine $\in$ as $\varepsilon^\flat$. This gives me the tools to reverse-engineer my symmetric theory, $\mathsf{LTU}_{\mathsf{sym}}$. I first define a direct translation, $\sharp$, as $\flat$'s inverse:

$$Set^\sharp(\mathrm{x}) :\equiv E(\mathrm{x},\mathrm{x}) \vee \exists v\ v\ \varepsilon\ \mathrm{x}$$
$$\mathrm{x} \in^\sharp \mathrm{y} :\equiv \mathrm{x}\ \varepsilon\ \mathrm{y}$$

Using this, I define $\mathsf{LTU}_{\mathsf{sym}}$ as the first-order theory with these axioms:

- $\phi^\sharp$, for every LTU-axiom ϕ
- $\forall y \forall y (E(x,y) \leftrightarrow E^{\flat\sharp}(x,y))$
- $\forall x \forall y (Tog(x,y) \leftrightarrow Tog^{\flat\sharp}(x,y))$

Our desired result now follows very easily:

Theorem D.4: $\mathsf{LTU}_{\mathsf{sym}}$ is symmetric and synonymous with LTU.

Proof. $\mathsf{LTU}_{\mathsf{sym}}$ is symmetric given its last two axioms; e.g. unpacking and simplifying $E^{\flat\sharp}(x,y)$ yields $x\ \varepsilon\ y \vee y\ \varepsilon\ x \vee (x = y \wedge E(x,x) \wedge \forall z\ z \not\varepsilon\ x)$. For synonymy: using Lemma D.1(1), note that $\mathsf{LTU} \vdash Set(a) \leftrightarrow Set^{\sharp\flat}(a)$; so via Lemma D.3, $\mathsf{LTU} \vdash \phi \leftrightarrow \phi^{\sharp\flat}$ for any LTU-formula ϕ. Now just invoke the construction of $\mathsf{LTU}_{\mathsf{sym}}$. □

References

Amijee, Fatema (2020). 'Principle of sufficient reason'. In: *The Routledge Handbook of Metaphysical Grounding*. New York: Routledge, pp.63–75.

Armstrong, David Malet (1978). *A Theory of Universals, Volume 2: Universals and Scientific Realism*. Cambridge: Cambridge University Press.

— (1997). *A World of States of Affairs*. Cambridge: Cambridge University Press.

Boolos, George et al. (2007). *Computability and Logic*. 5th ed. Cambridge: Cambridge University Press.

Button, Tim (2021). 'Level theory, part 1: Axiomatizing the bare idea of a cumulative hierarchy of sets'. *Bulletin of Symbolic Logic* 27.4, pp.436–60.

Button, Tim and Sean Walsh (2018). *Philosophy and Model Theory*. Oxford: Oxford University Press.

Carnap, Rudolf (1928). *Der logische Aufbau der Welt*. Berlin.

Castañeda, Héctor-Neri (1975). 'Relations and the identity of propositions'. *Philosophical Studies* 28.4, pp.237–44.

Cobham, Alan (1956). 'Reduction to a symmetric predicate'. *The Journal of Symbolic Logic* 21.1, pp.56–9.

Cross, Charles B (2002). 'Armstrong and the problem of converse relations'. *Erkenntnis* 56.2, pp.216–27.

Dipert, Randall R (1997). 'The mathematical structure of the world: The world as graph'. *The Journal of Philosophy* 94.7, pp.329–58.

Dixon, T Scott (2018). 'Plural slot theory'. *Oxford Studies in Metaphysics* 11, pp.192–223.

Dorr, Cian (2004). 'Non-symmetric relations'. *Oxford Studies in Metaphysics* 1, pp.155–192.

Fine, Kit (2000). 'Neutral relations'. *Philosophical Review* 109, pp.1–33.

Friedman, Harvey M and Albert Visser (2014). 'When bi-interpretability implies synonymy'. *Logic Group Preprint Series* 320, pp.1–19.

Hodges, Wilfrid (1993). *Model Theory*. Cambridge: Cambridge University Press.
Kaye, Richard and Tin Lok Wong (2007). 'On interpretations of arithmetic and set theory'. *Notre Dame Journal of Formal Logic* 48.4, pp.497–510.
Kemp, Gary (2017). 'Is everything a set? Quine and (Hyper)Pythagoreanism'. *The Monist* 100, pp.155–66.
Lakatos, Imre (1976). *Proofs and Refutations: The Logic of Mathematical Discovery*. Ed. by John Worrall and Elie Zahar. Cambridge: Cambridge University Press.
Lavrov, I A (1963). 'Effective inseparability of the set of identically true formulas and the set of formulas with finite counterexamples for certain elementary theories'. *Algebra i Logika* 2.1, pp.5–12.
Leo, Joop (2013). 'Relational complexes'. *Journal of Philosophical Logic* 42, pp.357–90.
Lewis, David (1983). 'New work for a theory of universals'. *Australasian Journal of Philosophy* 61.4, pp.343–77.
— (1986). *On the Plurality of Worlds*. Oxford: Blackwell.
Liebesman, David (2014). 'Relations and order-sensitivity'. *Metaphysica* 15.2, pp.409–29.
MacBride, Fraser (2007). 'Neutral relations revisited'. *Dialectica* 61.1, pp.25–56.
— (2014). 'How involved do you want to be in a non-symmetric relationship?' *Australasian Journal of Philosophy* 92, pp.1–16.
— (2015). 'On the origins of order: Non-symmetric or only symmetric relations?' In: *The Problem of Universals in Contemporary Philosophy*. Ed. by Gabriele Galluzzo and Michael J Loux, pp.173–94.
— (2020). 'Relations'. *Stanford Enyclopedia of Philosophy*. URL: https://plato.stanford.edu/archives/win2020/entries/relations/.
Marker, David (2000). *Model Theory: An Introduction*. London: Springer.
Orilia, Francesco (2014). 'Positions, ordering relations, and o-roles'. *Dialectica* 68.2, pp.283–303.
Quine, Willard van Orman (1954). 'Reduction to a dyadic predicate'. *The Journal of Symbolic Logic* 19.3, pp.180–2.
— (1976). 'Whither physical objects?' *Boston Studies in the Philosophy of Science* 39, pp.303–10.
Rabin, Michael O (1965). 'A simple method for undecidability proofs and some applications'. In: *Logic, methodology and philosophy of science, Proceedings of the 1964 International Congress*. Ed. by Y Bar-Hillel. Amsterdam: North-Holland, pp.58–68.
Russell, Bertrand (1903). *Principles of Mathematics*. London (2010): Routledge.
— (1913). *Theory of Knowledge: The 1913 Manuscript*. Ed. by E. R. Eames, in collaboration with K. Blackwell. London (1992): Routledge.
Sider, Theodore (2011). *Writing the Book of the World*. Oxford: Oxford University Press.
— (2020). *The Tools of Metaphysics and the Metaphysics of Science*. Oxford: Oxford University Press.
Svenonius, Lars (1955). 'Definability and simplicity'. *Journal of Symbolic Logic* 20.3, pp.235–50.
Szmielew, Wanda and Alfred Tarski (1950). 'Mutual interpretability of some essentially undecidable theories'. *Proceedings of the International Congress of Mathematicians* 1, p.734.
Tarski, Alfred (1954). 'A general theorem concerning the reduction of primitive notions'. *Journal of Symbolic Logic* 19.2, p.158.
Tarski, Alfred et al. (1953). *Undecidable Theories*. Amsterdam: North Holland.
Trueman, Robert (2021). *Properties and Propositions: The Metaphysics of Higher-Order Logic*. Cambridge: Cambridge University Press.
Visser, Albert (2008). 'Pairs, sets and sequences in first-order theories'. *Archive for Mathematical Logic* 47, pp.299–326.
Williamson, Timothy (1985). 'Converse relations'. *Philosophical Review* 94.2, pp.249–62.