A fictionalist theory of universals

Tim Button and Robert Trueman

September 18, 2023

We are Fregean realists. Very roughly speaking, this means that we believe in a vast type-hierarchy, and we insist that the typing is strict, so that every entity has a unique type. For example: we believe in properties, but we never confuse properties with objects.

Our question in this paper is whether Fregean realists should believe in universals as well as properties. By 'universals', we mean object-level correlates of properties, such as wisdom, mortality and the colour blue. There are good reasons to reject the existence of universals, but various natural language constructions appear to force us to believe in them. We explore a fictionalist response to this problem. Our fictionalist theory of universals allows us to speak as if universals existed, whilst denying that any really do.

We start by presenting our type theory in §1. Then, in §2, we introduce Fregean realism, and sketch the *Disquotation Argument* for it. In §3, we motivate a fictionalist account of universals. We present our particular brand of fictionalism in §§4–5, and apply it to a range of natural language constructions in §§6–7. We end by discussing the limits of our fictionalism in §8.

1 Partial-Functions Type Theory

In this paper, we will draw a sharp distinction between *properties* and *universals*. This might initially seem like a distinction without a difference; for us, however, it marks a crucial difference in type. We will operate with a version of Church's typed functional λ-calculus. The full details of the system are in §A, but we will start with a quick overview.

Our system has two basic types, `e` and `t`. Type `e` expressions correspond to natural language names, like 'Socrates' and 'Plato', and should be thought of as purporting to refer to objects.[1] Type `t` expressions correspond to natural language sentences, like 'Socrates is wise' and 'Plato pontificates', and should be thought of as expressing propositions.[2]

[1] Montague (1973: 18–19) treated natural language names as having type `(et)t`; Partee (1986: 360ff) instead suggests that they are 'basically of type `e`' and sometimes 'derivatively' lifted to type `(et)t`. We side with Partee, but everything we say in this paper could be reworked according to Montague's scheme.

[2] You might have thought that they express states of affairs rather than propositions, but see Trueman (2021: chs.11–13) for an argument that states of affairs *are* propositions.

We also have complex types: if α and β are types, then $(\alpha\beta)$ is also a type. We often omit outermost brackets in names for complex types, e.g. writing $\alpha(\alpha\beta)$ rather than $(\alpha(\alpha\beta))$.

A type $\alpha\beta$ expression combines with a type α expression to form a type β expression, as follows: $(\mathbf{B}^{\alpha\beta}\mathbf{A}^{\alpha})^{\beta}$. (If $\gamma \neq \alpha$, then $\mathbf{B}^{\alpha\beta}\mathbf{A}^{\gamma}$ is ill-formed.) Intuitively, an expression of type $\alpha\beta$ expresses a function from entities of type α to entities of type β. Throughout this paper, we move back and forth between types of *expression* and types of *entity*. As we understand it: an entity is of type α just in case it is a value of a type α variable. We reserve 'object' for entities of type $\mathtt{e}$.

Natural language predicates like 'is wise' and 'pontificates' are of type $\mathtt{et}$: they combine with names (type $\mathtt{e}$) to form sentences (type $\mathtt{t}$); 'Socrates' is a name, and when you combine it with 'is wise', you get the sentence 'Socrates is wise'. We call the functions that these predicates express *properties* of objects. In other words, properties of objects are functions of type $\mathtt{et}$, from objects to propositions. We also have other types of property in the type hierarchy. For example, type $(\mathtt{et})\mathtt{t}$ functions are properties of properties of objects. Intuitively, any function which has propositions as values is a property of some type. However, type $\mathtt{et}$ properties are our main focus in this paper, and they are the functions we mean by any unqualified use of 'property'.

Our theory includes the various logical constants you would expect, enabling us to handle sentential connectives and quantification over each type. We also have the device of λ-abstraction, and each type, α, has its own identity relation, $=_{\alpha}^{\alpha(\alpha\mathtt{t})}$. (In general: an expression's superscript indicates the expression's type, and we omit the superscript if it is obvious from context; an expression's subscript is an undetachable part of the expression, and it reminds us of a particularly salient argument-type to the expression. So in '$=_{\alpha}^{\alpha(\alpha\mathtt{t})}$', the subscript reminds us that this is an identity relation for type α entities.)

So far, our type theory is essentially Church's. We depart from Church in allowing for *partial functions*, and so we call our logic PFTT, for Partial-Functions Type Theory.[3] We have an explicitly defined existence predicate for each type, $\mathsf{E}_{\alpha}^{\alpha\mathtt{t}}$,[4] and a notion of 'identical if existent', defined as follows:

$$\mathbf{A}^{\alpha} \simeq_{\alpha} \mathbf{B}^{\alpha} := (\mathsf{E}\mathbf{A} \vee \mathsf{E}\mathbf{B}) \rightarrow \mathbf{A} =_{\alpha} \mathbf{B}$$

As this illustrates, we will use infix notation when convenient, and we will omit brackets, subscripts, and superscripts where doing so will improve readability and no ambiguity will arise. Likewise, whilst all functions in this system are monadic—so that polyadic functions are handled by currying—we often use uncurried notation for readability. For example, we let $\mathbf{B}^{\alpha_1(\alpha_2(\alpha_3\beta))}(\mathbf{A}_1^{\alpha_1}, \mathbf{A}^{\alpha_2}, \mathbf{A}^{\alpha_3})$ abbreviate $((\mathbf{B}^{\alpha_1(\alpha_2(\alpha_3\beta))}\mathbf{A}_1^{\alpha_1})\mathbf{A}_2^{\alpha_2})\mathbf{A}_3^{\alpha_3}$.

[3] Allowing for non-total functions allows us to deal easily with empty definite descriptions (see e.g. Heim and Kratzer 1998: 73–5, 154) or to handle claims like 'There is exactly one real number n such that $(3 \div n)$ does not exist' (example from Tichý 1988: 9).

[4] Given by $\mathsf{E}_{\alpha}\mathbf{A}^{\alpha} := \mathbf{A} =_{\alpha} \mathbf{A}$.

2 Fregean realism

Having outlined our type theory, we now want to discuss our philosophical attitude towards it. Simply put, we think that PFTT describes a hierarchy of types of entity, and that every entity has a unique type in this hierarchy. However, putting things this way requires a grain of salt. Our first aim in this section is to present our view, which we call *Fregean realism*, more carefully and without the seasoning (§2.1). After that, we will explain what we think motivates Fregean realism (§§2.2–2.3).

2.1 What is Fregean realism?

Fregean realism is a reaction against a more traditional form of realism. According to this traditional realism, *every* entity of *every* type is also an object; there is no type of entity which cannot (at least in principle) be named. Take properties, for example: properties are type `et` functions, which can be expressed by type `et` predicates, like 'is wise'; but according to traditional realism, properties are *also* type `e` objects, which can be referred to by type `e` names, like 'wisdom'.[5]

Fregean realists reject traditional realism. Roughly, Fregean realism is the doctrine that every entity has a unique type. So, for example, no property is an object:[6] the type `et` function expressed by 'is wise' cannot be referred to by any type `e` name, not even by 'wisdom'. (We call this 'Fregean' realism, because Frege famously insisted that no property—or in his terminology, no *concept*—is an object.)[7]

Unfortunately, though, that really is a *rough* statement of Fregean realism. The trouble is that, if every entity has a unique type, then it is impossible to say that every entity has a unique type. Focus on properties and objects again. If properties are not objects, then no function of type `et` can take a property as argument; in other words, nothing that can be said of an object can also be said of a property. We *can* say of an object that it is an object, and so 'is an object' must be a predicate of type `et`. (You might formalize it in PFTT as $(\lambda x^{\mathsf{e}} \exists y^{\mathsf{e}}\ x =_{\mathsf{e}} y)^{\mathsf{et}}$.) But that means that it must be *nonsense* to say of a property that it is an object. And since the negation of nonsense is also nonsense, it follows that we cannot say of a property that it is *not* an object. (And, indeed, both $(\lambda x^{\mathsf{e}} \exists y^{\mathsf{e}}\ x =_{\mathsf{e}} y)^{\mathsf{et}}\mathbf{A}^{\mathsf{et}}$ and its negation are ill-formed in PFTT.)

This is the heart of Frege's notorious concept *horse* paradox. To avoid this paradox, we must steadfastly avoid saying things like 'Properties are not objects'. We might

[5] Traditional realists include the early Russell (1903), Gaskin (1995, 2008), Wright (1998), Hale (2010, 2013: ch.1), Hale and Wright (2012), Hale and Linnebo (2020), MacBride (2011), Menzel (forthcoming: §1.4), Liebesman (2015), Price (2016), and Rieppel (2016, 2018).

[6] More generally, *no function is an object*, where 'function' is understood in the sense of PFTT. So, in particular, the functions of PFTT are not to be thought of as *sets*, or any other objects.

[7] Frege (1892). As many of the papers in this volume demonstrate, Fregean realism has enjoyed a recent surge of popularity (though not always by that name). Fregean realists include Prior (1971), Geach (1976), Rayo and Yablo (2001), Williamson (2003, 2013), Noonan (2006), Krämer (2014), Trueman (2015, 2021), Dorr (2016), Jones (2016, 2018), and Goodman (2017). (Prior would have resented being called a 'realist', but see Trueman (2021: chs 7 & 9) for discussion of just how minimal that label is in a higher-order setting.)

instead try to articulate Fregean realism as follows: different types of entity are incomparable, in the sense that what can be said of one type of entity cannot be said of another. That is certainly an improvement, but even this statement of Fregean realism is self-undermining. After all, to say that properties and objects are incomparable is still to try to compare them.

Really, then, Fregean realism should not be seen as a *doctrine* at all. Rather, Fregean realism is best seen as a kind of non-propositional attitude towards PFTT. To be a Fregean realist is to accept the limits of PFTT as the limits on what can be expressed, and so to dismiss any attempt to transcend them as nonsense: $\mathbf{B}^{\mathtt{et}} = \mathbf{A}^{\mathtt{e}}$ is ill-formed in PFTT, and so we dismiss any attempt to say that properties are, or are not, objects.[8] Still, though, it can often be convenient to speak *as if* Fregean realism were the doctrine that every entity has a unique type, and we will do so whenever that would not be too seriously misleading.

2.2 Against the abductive argument for Fregean realism

We are Fregean realists. We will give our reasons for adopting Fregean realism in §2.3. First, we would like to set aside a bad argument for Fregean realism.

A Fregean realist might try to motivate their view with a kind of abductive argument. The argument would go like this: Metaphysicians are aiming to find the metaphysical theory with the best balance of virtues (or perhaps just *a* metaphysical theory with a *sufficiently good* balance of virtues). Having the power to solve philosophical puzzles is high on the list of virtues for a metaphysical theory. And, happily, Fregean realism offers neat solutions to a range of longstanding puzzles. Here is one of the simplest (although not necessarily one of the deepest) examples:[9]

> *Platonists* believe that properties are not spatiotemporally located; *aristotelians* believe that properties are located when and where their instances are. However, it only makes sense to say that an *object* is or is not located: 'is located' is a predicate of type $\mathtt{et}$. So, given Fregean realism, it does not really make sense to say that a *property* is, or is not, located. If we tried, we would end up saying something nonsensical, like 'is wise is (not) located'. So, if Fregean realism is right, then the whole debate between platonists and aristotelians is misguided.

According to the abductive argument, solutions like this provide us with (defeasible) reasons to adopt Fregean realism.

We think that this abductive argument is extremely weak. Fregean realism does provide 'solutions' to several puzzles, but these 'solutions' will only satisfy those who *already* embrace Fregean realism. Every Fregean 'solution' to a puzzle is really a *dissolution*: it works by pointing out that any statement of the puzzle tries to say of a property something that can be said of an object, which is nonsensical by Fregean

[8] See Trueman (2021: ch. 9) for related discussion.

[9] Jones (2018), Trueman (2021: ch.10), and Skiba (2020) discuss this and other examples in more detail.

lights.[10] Such abduction-to-nonsensicality should have little sway with an impartial judge.

For one thing, the purported benefits of abduction-to-nonsensicality are not obviously benefits. To illustrate, return to the platonist/aristotelian debate. Fregean realists purport to resolve this debate by denying that it makes any sense. But, at least initially, the debate does *seem* to make sense: philosophers have certainly been arguing over it for some time, and, by and large, they *appear* to have understood one another. It is, then, not *obvious* that we should *want* our theory to dissolve the debate.

For another thing, the costs of abduction-to-nonsensicality can seem astronomical. The expressive limits imposed by Fregean realism do not just prevent us from formulating (for example) the platonism/aristotelianism debate; they also prevent us from formulating Fregean realism itself (see §2.1). Now, we have offered a way around the threat of self-stultification—by denying that Fregean realism should be thought of as a *doctrine*—but we recognise that many philosophers will consider this to be a serious cost of the view.[11]

Given all this, we doubt that an impartial judge should be won over to Fregean realism, if all they have to go on is the abductive argument. In fact, at this stage, they are much more likely to regard Fregean realists as trying to have their cake and eat it too: Fregean realists apparently help themselves to a rich ontology of functions, whilst attempting to excuse themselves from any deep metaphysical inquiry into that ontology on the (avowedly unintelligible) grounds that *functions are not objects*.

2.3 The Disquotation Argument for Fregean realism

Fortunately, there is a much better argument for Fregean realism, previously presented by Trueman.[12]

Suppose you want to say that some entities belong to more than one type. In particular, suppose you want to say that some property (type $\mathtt{et}$) is an object (type $\mathtt{e}$).[13] As we noted in §2.1, that is not something that can be said within PFTT: there is no reading of '=' on which $\mathbf{B}^{\mathtt{et}} = \mathbf{A}^{\mathtt{e}}$ is well-formed. So to say that some property is an object, you will have to find some way of transcending PFTT. But there are only two strategies for trying to do that, and both fail.

[10] Not that this is always obvious. For example, Williamson (2003) has shown that Fregean realists can account for absolutely unrestricted first-order quantification without running into Russell's Paradox. On the face of it, this has nothing to do with denying that certain claims make sense. However, Fregean realists can only accommodate unrestricted quantification because they deny that it makes sense to ask whether a domain could include both properties and objects (see Button and Trueman 2022: §7, *contra* Florio and Jones 2019).

[11] For example, see Linnebo (2006: §.4), Hale and Wright (2012: §III), Proops (2013), and Hale and Linnebo (2020).

[12] Trueman (2021: chs 1–9).

[13] We focus on this case, but a similar argument will work no matter what types you choose.

Strategy 1: relaxing the formation rules. In PFTT, names and predicates are never intersubstitutable *salva congruitate*: if $\phi(\mathbf{A}^{\mathtt{e}})$ is a well-formed sentence, then $\phi(\mathbf{B}^{\mathtt{et}})$ is not. But we could relax those formation rules, and allow names and predicates to be intersubstitutable in some, or even *all*, contexts.[14] If we did, then there would be nothing stopping us from admitting $\mathbf{B}^{\mathtt{et}} = \mathbf{A}^{\mathtt{e}}$ as a well-formed formula.

However, it is important to remember that the formation rules built into PFTT are not just arbitrary syntactic impositions. Type `e` expressions are meant to correspond to natural language names like 'Socrates' and 'Plato'; and type `et` expressions are meant to correspond to natural language predicates like 'is wise' and 'pontificates'. Names and predicates play two very different semantic roles. Consider the following natural language sentence, along with its formalization in PFTT:

(1) Socrates is wise
$wise^{\mathtt{et}}(socrates^{\mathtt{e}})$

In this sentence, 'Socrates' refers to Socrates, and 'is wise' says of him that he is wise. More generally, names *refer* to objects, and predicates *say things of* objects. (That is the sense in which predicates express functions from objects to propositions.) These roles are clearly designed to work together, and if we try to intersubstitute them, we end up with meaningless nonsense, such as:

(2) Socrates Plato
$plato^{\mathtt{e}}(socrates^{\mathtt{e}})$

(3) pontificates is wise
$wise^{\mathtt{et}}(pontificates^{\mathtt{et}})$

Crucially, (2) is not just *ungrammatical*, but wholly *meaningless*: 'Plato' is a name, and so its job is merely to refer to an object, not to say anything of the referent of 'Socrates'.[15] Similar remarks apply to (3): 'pontificates' is a predicate, and so its job is to say something of an object, not merely to provide a referent for 'is wise' to say something of.[16]

It is not an option, then, simply to relax PFTT's formation rules. If type `e` expressions behave as names, and type `et` expressions behave as predicates, then it would be meaningless to intersubstitute them.

Strategy 2: metalinguistic ascent. Rather than trying to say that some property is an object directly in PFTT, we might try to say it indirectly in a metalanguage. For example, here is how we might try to say that property $b^{\mathtt{et}}$ is identical to object $a^{\mathtt{e}}$:

[14] Linnebo and Rayo's (2012) *cumulative* type theory relaxes its formation rules in just this way; for criticism, see Button and Trueman (2022).

[15] Magidor (2009) claims that (2) is meaningful but false; we think this overlooks the crucial difference in semantic role between names and predicates.

[16] This argument does not *assume* that 'pontificates' is non-referring. (That assumption would be question-begging in the current context.) The point is that, even if 'pontificates' refers to an object, it is *also* meant to say something of an object, which is a role it cannot discharge in (3).

(4) the referent of '$b^{\mathtt{et}}$' = the referent of '$a^{\mathtt{e}}$'

But why should we think of (4) as an indirect way of identifying $b^{\mathtt{et}}$ with $a^{\mathtt{e}}$? The answer must be that we are presupposing that $b^{\mathtt{et}}$ is the referent of '$b^{\mathtt{et}}$', and $a^{\mathtt{e}}$ is the referent of '$a^{\mathtt{e}}$'. In other words, we are presupposing that we can use reference to disquote '$b^{\mathtt{et}}$' and '$a^{\mathtt{e}}$'.[17] However, since (as we have just argued) names and predicates cannot be meaningfully intersubstituted, no single notion of reference could be used to disquote predicates as well as names. Instead, we will need two *referent* functions, one of type $\mathtt{ee}$ to disquote names, and one of type $\mathtt{e(et)}$ to disquote predicates:[18]

(5) $\mathit{referent}^{\mathtt{ee}}_{\mathtt{e}}(\text{`}a^{\mathtt{e}}\text{'}) =_{\mathtt{e}} a^{\mathtt{e}}$
(6) $\mathit{referent}^{\mathtt{e(et)}}_{\mathtt{et}}(\text{`}b^{\mathtt{et}}\text{'}) =_{\mathtt{et}} b^{\mathtt{et}}$

And now that we have drawn this distinction between $\mathit{referent}_{\mathtt{e}}$ and $\mathit{referent}_{\mathtt{et}}$, we face the same problem in the metalanguage that we previously faced in the object-language: '$\mathit{referent}_{\mathtt{e}}(\text{“}a^{\mathtt{e}}\text{”})$' and '$\mathit{referent}_{\mathtt{et}}(\text{“}b^{\mathtt{et}}\text{”})$' have different types—respectively $\mathtt{e}$ and $\mathtt{et}$—so that '$\mathit{referent}_{\mathtt{e}}(\text{“}a^{\mathtt{e}}\text{”}) = \mathit{referent}_{\mathtt{et}}(\text{“}b^{\mathtt{et}}\text{”})$' is ill-formed.[19] (Of course, you might try to sidestep this problem by relaxing the formation rules in your metalanguage, but that would just be a re-run of Strategy 1.)

That, in a nutshell, is the *Disquotation Argument* for Fregean realism. We should emphasise that this is a quick summary of a complex argument, presented fully elsewhere (Trueman 2021: chs 1–9). There are a number of points at which you might object. Ultimately, though, we think that this argument succeeds, and this is why we are Fregean realists.

3 Universals and Nominalization

Although we are Fregean realists, we must admit that natural languages *appear* to flout Fregean realism's strict type-distinctions. In particular, natural languages provide us with a variety of devices for *nominalizing* predicates, i.e. for converting predicates (expressions of type $\mathtt{et}$) into names (expressions of type $\mathtt{e}$). To illustrate, consider these two natural English sentences:

(1) Socrates is wise
(7) Wisdom is a virtue

[17] We are not assuming that reference is *exhausted* by disquotational principles; only that disquotational principles are *correct* (for the home language).

[18] As in §1: the subscripts on '$\mathit{referent}_{\mathtt{e}}$' and '$\mathit{referent}_{\mathtt{et}}$' are undetachable; they serve as mnemonics for the types of expression they can disquote.

[19] Equally, '$(\lambda x^{\mathtt{e}} \exists y^{\mathtt{e}}\ x =_{\mathtt{e}} y)^{\mathtt{et}}(\mathit{referent}_{\mathtt{et}}(\text{“}b^{\mathtt{et}}\text{”}))$' is also ill-formed.

'Wisdom' is a nominalization of 'wise'.[20] But intuitively, despite this type difference, the name 'wisdom' should refer to the very property expressed by the predicate 'wise': what (7) declares to be a virtue should be precisely the property that (1) applies to Socrates.

What is more, nominalization appears to be a feature, not a bug, of natural language. Anyone who has spent any time working within a strictly typed system will know just how *difficult* it can be to obey type-restrictions consistently. (To give just one example, in a strictly typed system, it is impossible to generalize over every type of entity all at once; at best, we can *simulate* such generalizations with *typically ambiguous* schemes.) Pushing every entity down into type `e`, where old type distinctions can safely be ignored, makes natural language a far more convenient communicative tool.

The main aim of this paper is to resolve this tension between Fregean realism and natural language. In this section, we will sketch (§3.1) and motivate (§§3.2–3.4) our preferred resolution.

3.1 Fictionalism about universals

The first thing to be clear about is that Fregean realism does not forbid nominalization. All that Fregean realism tells us is that nominalized predicates, being type `e` names, cannot refer to type `et` properties. At best, they can refer to special, object-level correlates of properties, which we will call *universals*. (This is stipulative but not unmotivated: the argument we are about to present is one of Armstrong's arguments for universals.[21]) Given Fregean realism, it would be a type-confusion to identify a property with a universal.

In our new terminology, then, nominalized predicates are names which purport to refer to universals. But do they actually succeed? Are there any universals, or are nominalized predicates systematically *empty* names? A familiar style of argument seems to show that some nominalized predicates do, indeed, successfully refer to universals. Return to (7):

(7) Wisdom is a virtue

Intuitively, this sentence is not just meaningful, but also *true*.[22] And given standard semantic clauses, (7) cannot be true unless 'wisdom' is a referring name. So 'wisdom' refers, and at least one universal exists.[23]

[20] In principle, you could try to dispute this. In particular, you could try claiming that 'wisdom' is really type `et`, just like 'wise'. However, the cases of mixed-predication discussed in §7.1—e.g. 'Plato loves Socrates and wisdom'—provide clear linguistic evidence that nominalized predicates are genuine names. This point is rightly emphasized by Hofweber (2018: §3).

[21] Armstrong (1978: ch. 6).

[22] If you disagree that (7) is intuitively true, substitute in your favourite example.

[23] Closely related arguments have been presented by: Pap (1959), Jackson (1977), Armstrong (1978: ch. 6), Schiffer (2003: §2.3), and Thomasson (2014: ch. 3). There is, of course, another famous argument for universals, namely the *Problem of Universals* (see Armstrong 1978, 1980, 1989: 88–9, 2004: 39–42;

We want to resist this argument, and so we deny that (7) is really true. More generally, we deny that *any* atomic sentence featuring a nominalized predicate is true. However, we do not wish to deny that many of these sentences are still *assertible*. Instead, we advocate a *fictionalism about universals*, according to which a sentence about universals is assertible iff it is *true in the fiction of universals*. We lay out the details of this fictionalism in §§4–5. First, though, we will present three reasons why a Fregean realist should be a fictionalist about universals.[24]

3.2 Motivating fictionalism: representational aids

Here is a natural story for Fregean realists to tell about the usefulness of nominalization:

> As we conceded earlier, actually speaking in a strictly typed language is often inconvenient. Indeed, even the most ardent Fregean realists usually allow themselves to speak 'loosely', as if properties were objects. But what we are really doing when we speak 'loosely' is introducing objects to *represent* real properties. These are the objects that we are now calling 'universals'. These universals are really nothing more than representational aids: it is worth asserting sentences about universals because those sentences have implications about properties, and it is those implications that really matter.

We think that this story is basically right. And it would be a small step from here to a fictionalism about universals.

Granted, it would still be *a step*. You could consistently concede that universals are just representational aids, whilst insisting that they really exist.[25] But fictionalism seems like the more attractive option (assuming its details can be worked out). If the whole value of an assertion about universals is its implications for properties, it would seem gratuitous to postulate any special entities just to make that assertion true.[26]

3.3 Motivating fictionalism: philosophical puzzles

In §2.2, we noted that Fregean realism promises to solve various philosophical puzzles about properties. Now, to repeat: we do not think that these solutions by themselves provide any motivation for Fregean realism; rather, Fregean realism is motivated by

Rodriguez-Pereyra 2000). We will not consider that argument here, since Trueman (2021: §10.1) has already argued that the Problem of Universals has no force against Fregean realists.

[24] Bâve (2015) also recommends combining higher-order logic with fictionalism about universals. However, he does not provide a Conservativeness Theorem (contrast our §§4.2 & C), and he does not discuss mixed-predication (contrast our §§7.1 & E). For further discussion of Bâve, see footnote 30.

[25] In fact, this seems to have been Frege's (1892) view at one time. By way of contrast, see Frege (1891–5, 1924/5: 269–70).

[26] This motivation for fictionalism about universals is structurally identical to one of the standard motivations for fictionalism about mathematical entities (see Field 1980/2016; Balaguer 1996; Yablo 2005). However, that does not mean that these two fictionalisms stand or fall together. It might be, for example, that fictionalism about mathematical entities faces special difficulties that do not confront our fictionalism about universals.

the Disquotation Argument of §2.3. Nonetheless, once you have been convinced by the Disquotation Argument, it is philosophically pleasing that you can now (dis)solve these puzzles.

However, if we admit universals into our ontology, then all those old puzzles will threaten to return as puzzles about universals. As we explained in §2.2, the Fregean solutions work by observing that, since properties are not objects, the traditional puzzles about properties are really just nonsensical pseudo-problems. But universals *are* objects. So Fregean realists could not so easily dismiss these problems, if they were reworked to concern universals. For example, the debate between platonists and aristotelians would start back up, this time as a debate over whether *universals* are spatiotemporally located.

Now, that would not be a *complete* disaster. Even if all the old puzzles did return as puzzles about universals, we would still have made progress by demonstrating that they do not concern the notion of *property* involved in predication.[27] Nonetheless, Fregean realism undeniably gives us the most philosophical bang for our philosophical buck if we deny that there are any universals.

3.4 Motivating fictionalism: Cantor's Theorem

Any systematic theory of universals needs a device of nominalization.[28] Intuitively, given an input property, this device should output the corresponding universal. For now, we will represent this device by underlining, so that $\underline{wise}$ is wisdom. Now, consider two attractive principles:

Nom-Coext. $\forall u^{\mathsf{et}} \forall v^{\mathsf{et}} (\underline{u} =_{\mathsf{e}} \underline{v} \rightarrow \forall x^{\mathsf{e}} (ux \leftrightarrow vx))$
Nom-Always. $\forall u^{\mathsf{et}} \, \mathrm{E}\underline{u}$

Principle Nom-Coext says that properties which correspond to the same universal are coextensive. This allows that some properties may not correspond to any universal. But that is ruled out by Nom-Always, which says that *every* property corresponds to a universal.

Unfortunately combining Nom-Coext with Nom-Always immediately leads to inconsistency, by an unsurprising version of Cantor's Theorem.[29] We must, then, choose between Nom-Coext and Nom-Always.[30]

[27] In fact, the Fregean realist solution to Bradley's Regress works whether or not there are universals; it requires only that we not identify universals with properties. See Trueman (2021: §10.2) for details.

[28] For other uses of a nominalization operator, see e.g. Cocchiarella (1974: 552–3), Chierchia (1984: 47ff, 1985: 422ff), Chierchia and Turner (1988: 266ff), Partee (1986: 362–3), and Hale and Linnebo (2020: 86ff).

[29] See Corollary 3 in §A.3. This holds in PFTT, so the problem could in principle be blocked by (substantially) revising PFTT (e.g. by embracing paraconsistency, or by modifying Comprehension). We do not have space to explore alternative logics, but we remark that the cost is high: PFTT seems perfectly suited for Fregean realists, up until the moment we are forced to consider universals.

[30] In footnote 24, we mentioned Båve's fictionalism about universals. In our terms, Båve's (2015: 29) fiction includes Nom-Always and $\forall x^{\mathsf{e}} \forall u^{\mathsf{et}} ((x \text{ instantiates } \underline{u}) \leftrightarrow ux)$. Given standard comprehension,

The choice is easy, and independent of our advocacy of Fregean realism. If Nom-Coext fails, then there are a^{et} and b^{et}, such that a-ness is b-ness, even though some object is a but not b, i.e. some x^{e} is such that ax but $\neg bx$. Since ax but $\neg bx$, presumably also x instantiates a-ness but not b-ness. But, despite all this, we are supposed to insist that a-ness is b-ness. That is surely absurd.[31] We therefore embrace Nom-Coext.

Accordingly, we must reject Nom-Always, and allow that some properties correspond to no universal. Indeed, since we are dealing with Cantor's Theorem, *most* properties correspond to no universal.[32] So we face an obligatory question: *which properties have corresponding universals?* Fictionalists offer two pleasingly simple answers:

Literally speaking: no property has a corresponding universal.
Fictionally speaking: all and only the real properties (i.e. the properties which exist, literally speaking) have corresponding universals.

The literal answer follows immediately from the fictionalist's (literal) denial that there are any universals. The fictional answer is just a well-motivated principle which we can (and will) embrace when setting up our fiction (see §4).

This provides our final motivation for fictionalism about universals: realists about universals have no similarly simple answer to the obligatory question. To begin with, they cannot mirror the fictionalist's fictional answer. After all, being realists, they draw no real/fictional distinction. So if they try to say that, literally, all and only the real properties have corresponding universals, this amounts to the claim that *every* property has a corresponding universal; and that is precisely the (catastrophic) principle Nom-Always.[33] Realists about universals cannot, then, treat all real properties equally: they can only allow a select few to correspond with universals. But it will still be useful to speak as if all real properties corresponded to universals, for the reasons given in §3.2. So, even would-be realists should be fictionalists about most universals. And, at this point, it is not clear what there is to be gained from resisting our thoroughgoing fictionalism, which applies to all universals across the board.

these are inconsistent: just consider $\lambda x^{\mathsf{e}} \neg (x \text{ instantiates } x)$. Båve gives no suggestion as to how he would avoid this inconsistency.

[31] Although one person's ponens is another's tollens: Cocchiarella (1972: 169, 1974, 1975a: 41–2, 1975b: 346–7) developed a system which can be (re-)interpreted as rejecting Nom-Coext in favour of Nom-Always.

[32] *Pace* Partee's (1986: 363) claim that nominalization 'is "almost" total'.

[33] Some pragmatist-inclined philosophers accuse fictionalists of drawing a spurious distinction between the real and the fictional (Schiller 1912: 99–100; Blackburn 1987: 56–60, 2005; Thomasson 2013: 1039, 2014: 197). In general, we think that this is an important challenge against fictionalists. However, the reasoning we just offered gives us a good answer to it in this particular case: collapsing the real/fictional distinction would plunge us right back into the tricky predicament we just raised against realists about universals.

4 A formal theory of universals

We have outlined our reasons for favouring fictionalism about universals. Our aim in this section is to provide a formal theory which can handle universals. In the next section, we will develop our fictionalist interpretation of that formal theory.

4.1 Restricting to real entities

Let T be a PFTT-theory which we use to assert sober, literal truths, without making any mention of universals. So, T will not include any nominalization operators. The idea is to move from T to some *fictional* theory, T_{U}, which does admit universals.

We start by modifying T itself. Implicitly, T talks only about real entities (i.e. entities that literally exist); this needs to be made explicit. We do this by introducing new constants, $real^{\alpha\mathsf{t}}_{\alpha}$, for each α, reading '$real^{\alpha\mathsf{t}}_{\alpha}(\mathbf{A}^{\alpha})$' as '$\mathbf{A}$ is real'. (In what follows, we tend to suppress both subscripts and superscripts on '*real*', since they are obvious from context.) We then replace each sentence $\mathbf{A}^{\mathsf{t}}$ of T with $\mathbf{A}^{\mathsf{t}}_{\text{R}}$, where the latter results from the former by restricting all discourse to real entities. If we were simply using first-order logic, we could achieve this by mapping each formula $\forall x \phi$ to $\forall x (real(x) \rightarrow \phi)$. Since PFTT is a much richer logic, the specification of the restriction is inevitably more complicated, but the intuitive idea is just the same (see §B for the details). The result is the real-restricted version of T, the theory $T_{\text{R}} = \{\mathbf{A}^{\mathsf{t}}_{\text{R}} : T \vdash \mathbf{A}^{\mathsf{t}}\}$.

4.2 Nominalization and Application Theory, NAT

The next step is to augment T_{R} with a theory which describes the behaviour of nominalization in general. We call this NAT, for Nominalization and Application Theory. To spell it out, we begin with some intuitive ideas.

Nominalization. We need devices for nominalizing (real) higher-typed entities. To this end, we have a constant, $nom^{\alpha\mathsf{e}}_{\alpha}$, for any type $\alpha \neq \mathsf{e}$.[34] For readability, we extend the convention developed in §3.4 and abbreviate $nom_{\alpha}(\mathbf{A}^{\alpha})$ as $\underline{\mathbf{A}^{\alpha}}$. (So, we continue to regard wisdom as $\underline{wise}$.) Mnemonically, think of underlining as pulling an entity down to the level of objects. To make certain principles easier to formulate, we sometimes also write $\underline{\mathbf{A}^{\mathsf{e}}}$ as an alternative to $\mathbf{A}^{\mathsf{e}}$, even though there is no operator nom_{e}, and no need for one.

Application. In PFTT, we apply higher-order entities to one another. We want a way to keep track of this among their nominalizations.[35] So, for each n, we intro-

[34] Natural languages actually allow us to nominalize expressions in a variety of different ways, and you might think that different kinds of nominalization refer to different kinds of object. For example, you might think that the gerund 'Sharon's laughing' refers to an event, whereas the clause 'that Sharon laughs' refers to a reified-proposition (see §6.4). Our focus in this paper is exclusively on nominalizations which intuitively appear to co-refer with their de-nominalized counterparts (cf. the argument at the start of §3), and this focus is reflected in the axioms of NAT.

[35] Strictly speaking, nominalization is a language-level operation, but for ease of expression, we will also use 'nominalization' to describe the corresponding world-level operation; so we will call the name

duce a new constant, $app_n^{\mathsf{e}(\ldots(\mathsf{ee})\ldots)}$, with n+2-occurrences of e, i.e. it will map n+1 input objects to an object. We read '$app_n(\mathbf{B}, \mathbf{A}_1, \ldots, \mathbf{A}_n)$' as 'the result of applying $\mathbf{B}$ to $\mathbf{A}_1, \ldots, \mathbf{A}_n$ in that order'. So $app_2(\underline{loves}^{\mathsf{e}(\mathsf{et})}, plato, socrates)$ is the result of applying Love to Plato and Socrates (in that order), and we want this to be identical$_{\mathsf{e}}$ to $\underline{loves}^{\mathsf{e}(\mathsf{et})}(plato, socrates)$.

We can now lay down the axiom schemes of NAT:[36]

Nom-real	$real(x^{\alpha}) =_{\mathsf{t}} \mathsf{E}\underline{x}$	all $\alpha \neq \mathsf{e}$
Nom-nonreal	$\neg real_{\mathsf{e}}(\underline{v^{\alpha}})$	all $\alpha \neq \mathsf{e}$
Prop-real	$real(x^{\mathsf{t}})$	
Nom-inj	$(\mathsf{E}\underline{u^{\alpha}} \wedge \mathsf{E}\underline{v^{\alpha}}) \rightarrow (u =_{\alpha} v) =_{\mathsf{t}} (\underline{u} =_{\mathsf{e}} \underline{v})$	all $\alpha \neq \mathsf{e}$
Nom-diff	$\underline{u^{\alpha}} \neq_{\mathsf{e}} \underline{v^{\beta}}$	$\alpha, \beta, \mathsf{e}$ all distinct
Application	$(\mathsf{E}\underline{u^{\alpha_1(\alpha_2(\ldots(\alpha_n\beta)\ldots))}} \wedge \mathsf{E}\underline{v_1^{\alpha_1}} \wedge \ldots \wedge \mathsf{E}\underline{v_n^{\alpha_n}}) \rightarrow$ $app_n(\underline{u}, \underline{v_1}, \ldots, \underline{v_n}) \simeq_{\mathsf{e}} \underline{u(v_1, \ldots, v_n)}$	all $\alpha_1, \ldots, \alpha_n, \beta$

The Nom-real scheme tells us that all and only real entities have nominalizations, and Nom-nonreal says that nominalizations are never real. However, all propositions are real, by Prop-real. Next, Nom-inj and Nom-diff schemes tell us that nominalization is injective, i.e. that nominalizations of entities are identical iff those entities are identical.[37] Finally, the Application scheme says that application among nominalized entities tracks the behaviour of higher-typed entities.

This completes the specification of NAT. We now define T_{U} as $T_{\mathrm{R}} \cup$ NAT, i.e. the addition of NAT to T, when the latter's implicit restriction to real entities is made explicit. Our central result is that T_{U} is a *conservative* extension of T, in this sense (see §C):

Conservativeness Theorem: Let T be a PFTT theory in some signature, $\mathscr{L}$, which is disjoint from NAT's non-logical vocabulary. If $T_{\mathrm{U}} \vdash \mathbf{A}_{\mathrm{R}}$ then $T \vdash \mathbf{A}$, for any $\mathscr{L}$-sentence $\mathbf{A}^{\mathsf{t}}$.

'$\underline{wise}$' a nominalization of the predicate '*wise*', but we will also call the *object* $\underline{wise}$ a nominalization of the *property wise*.

[36] Here, and throughout, $\mathbf{A} \neq_{\alpha} \mathbf{B}$ abbreviates $\neg(\mathbf{A} =_{\alpha} \mathbf{B})$. NAT bears some similarities to Hale and Linnebo's (2020: 102–3) theory of nominalization. Compare our Nom-inj with their (Bridge-=); and our Application with their (Bridge-*App*). The main technical differences concern: our use of PFTT (Hale and Linnebo use monadic relational type theory); our desire that T_{U} should obey unrestricted Comprehension (contrast Hale and Linnebo 2020: 103n.53); and our subsequent inclusion of bridge-principles which will ultimately allow for self-predication (see §7.3). Note also that Hale and Linnebo are traditional realists.

[37] In fact, Nom-inj gives us more than the biconditional $(u =_{\alpha} v) \leftrightarrow \underline{u} =_{\mathsf{e}} \underline{v})$; it gives us a propositional-identity $(u =_{\alpha} v) =_{\mathsf{t}} (\underline{u} =_{\mathsf{e}} \underline{v})$. Certain realists about universals might favour the weaker principle. But we are fictionalists, aiming for the Conservativeness Theorem; and the conservativeness of the stronger principle immediately entails the conservativeness of the weaker principle. (Similar remarks apply to Nom-real.) That said, our definition of instantiation relies explicitly on identity; see §4.3 and footnote 42.

This Theorem immediately entails NAT's consistency.[38] But it is worth specifically noting that NAT avoids the inconsistency we discussed in §3.4 by disavowing Nom-Always (importantly, Nom-inj is a conditional). Moreover, as promised, NAT proves Nom-Coext,[39] and states that all and only the real properties correspond to universals (via Nom-real).

4.3 Definable notions and richer fictions

Within NAT, we can define some further notions, which will be extremely useful in what follows.

Flattening. Any type $\alpha_1(\ldots(\alpha_n\mathtt{t})\ldots)$ property can be *flattened*. We write the flattening of **A** as $\underset{\sim}{\mathbf{A}}$. Intuitively, some objects satisfy $\underset{\sim}{\mathbf{A}}$ iff those objects are the nominalizations of entities that satisfy **A**. In detail, for each $\mathbf{A}^{\alpha_1(\ldots(\alpha_n\mathtt{t})\ldots)}$:[40]

$$\underset{\sim}{\mathbf{A}}(\underline{v_1^{\alpha_1}},\ldots,\underline{v_n^{\alpha_n}}) =_{\mathtt{t}} \mathbf{A}(v_1,\ldots,v_n) \qquad \text{if } \exists\underline{v_i} \text{ for all } 1 \leq i \leq n$$
$$\underset{\sim}{\mathbf{A}}(x_1,\ldots,x_n) =_{\mathtt{t}} \mathit{False}_{\mathtt{t}} \qquad \text{for all other cases}$$

To illustrate, PFTT allows us to symbolize the claim that something is wise as $\Sigma_{\mathtt{e}}(\mathit{wise}^{\mathtt{et}})$; $\Sigma_{\mathtt{e}}(\mathit{wise})$ is then identical$_{\mathtt{t}}$ to $\underset{\sim}{\Sigma}_{\mathtt{e}}(\underline{\mathit{wise}})$.

Instantiation. We can also define a notion of instantiation. Specifically, we can define a formula $\mathbf{A}\ \varepsilon\ \mathbf{B}$, which should be read '**A** instantiates **B**'. Thus, 'Socrates instantiates wisdom' can be formalized as '$\mathit{socrates}^{\mathtt{e}}\ \varepsilon\ \underline{\mathit{wise}^{\mathtt{et}}}$'. More generally, for each n, we can explicitly define an n+1-place instantiation relation, $\varepsilon_n^{\mathtt{e}(\ldots(\mathtt{et})\ldots)}$, as follows (we drop the subscript on 'ε_n', when no confusion can arise):[41]

$$(y_1^{\mathtt{e}},\ldots,y_n^{\mathtt{e}}\ \varepsilon_n\ x^{\mathtt{e}}) := \begin{cases} s^{\mathtt{t}} & \text{if } \underline{s} =_{\mathtt{e}} \mathit{app}_n(x,y_1,\ldots,y_n) \\ \mathit{False}_{\mathtt{t}} & \text{if there is no such } s^{\mathtt{t}} \end{cases}$$

We might pronounce the left-hand-side here as '$y_1,\ldots,y_n$, in that order, instantiate x'. So, to say that $(\mathit{plato}, \mathit{socrates}\ \varepsilon_2\ \underline{\mathit{loves}^{\mathtt{e}(\mathtt{et})}})$ is to say that Plato and Socrates, in that order, instantiate Love. Now NAT proves the following scheme, which is analogous to Application:[42]

---

[38] This is because an inconsistent theory is conservative only over inconsistent theories. More formally: let T be the PFTT theory with no axioms; suppose NAT is inconsistent, i.e. NAT $\vdash \bot$; then $T_{\mathrm{U}} \vdash \bot$, so, by the Conservativeness Theorem, $T \vdash \bot$, which is absurd.

[39] Indeed, it proves the stronger principle $\underline{u^{\mathtt{et}}} =_{\mathtt{e}} \underline{v^{\mathtt{et}}} \rightarrow u =_{\mathtt{et}} v$. *Proof.* Suppose $\underline{u^{\mathtt{et}}} = \underline{v^{\mathtt{et}}}$. So $\exists\underline{u}$ and $\exists\underline{v}$, by $Cr_{\mathtt{t}}$ (see §A.3); now $u =_{\mathtt{et}} v$ by Nom-inj and our initial supposition.

[40] At the expense of legibility, this definition can be presented austerely in PFTT. For example, let D abbreviate $(\imath x^{\alpha} v^{\mathtt{e}} =_{\mathtt{e}} \underline{x})$; then $\underset{\sim}{\mathbf{A}}^{\alpha\mathtt{t}}$ is $\lambda v^{\mathtt{e}} \imath s^{\mathtt{t}}((\exists\mathrm{D} \rightarrow s =_{\mathtt{t}} \mathbf{A}\mathrm{D}) \wedge (\neg\exists\mathrm{D} \rightarrow s =_{\mathtt{t}} \mathit{False}_{\mathtt{t}}))$.

[41] At the expense of legibility, this definition can be presented austerely in PFTT. For example, let A abbreviate $\mathit{app}_1(x,y)$; then ε_1 is $\lambda y^{\mathtt{e}} \lambda x^{\mathtt{e}} \imath s^{\mathtt{t}}((\exists\mathrm{A} \rightarrow \underline{s} =_{\mathtt{e}} \mathrm{A}) \wedge (\neg\exists\mathrm{A} \rightarrow s =_{\mathtt{t}} \mathit{False}_{\mathtt{t}}))$.

[42] *Proof.* By Prop-real and Nom-real, $\exists(\underline{u(v_1,\ldots,v_n)})$. So $\mathit{app}_n(\underline{u},\underline{v_1},\ldots,\underline{v_n}) =_{\mathtt{e}} \underline{u(v_1,\ldots,v_n)}$ by Application. Now use Nom-inj and the definition of ε_n. As mentioned in footnote 37, this (essentially) relies upon the propositional-identity in Nom-inj, which some realists about universals might reject. Indeed, they are likely to eschew the identification, $(\underline{v_1},\ldots,\underline{v_n}\ \varepsilon_n\ \underline{u}) =_{\mathtt{t}} u(v_1,\ldots,v_n)$, in favour of the

$$(\mathsf{E}\underline{u}^{\alpha_1(\alpha_2(\ldots(\alpha_n\mathsf{t})\ldots))} \wedge \mathsf{E}\underline{v_1}^{\alpha_1} \wedge \ldots \wedge \mathsf{E}\underline{v_n}^{\alpha_n}) \rightarrow$$
$$(\underline{v_1}, \ldots, \underline{v_n}\ \varepsilon_n\ \underline{u}) =_{\mathsf{t}} u(v_1, \ldots, v_n) \qquad \text{for all } \alpha_1, \ldots, \alpha_n$$

So: $(socrates\ \varepsilon\ \underline{wise})$ is identical$_{\mathsf{t}}$ to $wise(socrates)$.

For all its strengths, the theory NAT is very far from complete. To illustrate: NAT is compatible with the claim that $(plato^{\mathsf{e}}\ \varepsilon\ socrates^{\mathsf{e}})$, which is the sort of thing we might want to rule out, since no one, surely, wants to say that Plato *instantiates* Socrates. We can easily enrich NAT to rule out such things; the result remains conservative; and indeed NAT can be enriched further still, whilst retaining the Conservativeness Theorem. We say more about this in §7 and §D, but we will not dwell on it now since, for the time being, we will only need the principles mentioned in §4.2.

5 Fictionalism

We think that NAT is literally false. In fact, we can locate its falsity very precisely: five of NAT's six schemes are true; but Nom-real is false. After all, since no universals exist, '$\mathsf{E}(\underline{wise})$' is false, but '$real(wise)$' is true, so that '$real(wise) =_{\mathsf{t}} \mathsf{E}(wise)$' is false.

But although NAT is literally false, it is still a *useful fiction*. Specifically, and as we will now explain, our Conservativeness Theorem allows us to advance a fictionalism about universals, modelled after Field's fictionalism about numbers.[43]

Our fictionalism is primarily intended to answer the following obligatory question: *Why is it that, barring occasional mistakes by individual speakers, speaking as if universals existed does not lead us from obvious truths to obvious falsehoods?* Realists about universals have an easy answer to this question: *because universals really do exist*. But our Conservativeness Theorem provides us with an alternative answer: *even though there are no universals, positing such universals is conservative over the literal truths, and this is why they never lead us astray.*

Here is that answer in a little more detail. Suppose we start with some PFTT theory, T. We then make explicit the implicit assumption that T concerns only real entities, and so move to T_{R}. Now, by the Conservativeness Theorem, any *real-restricted claim* (i.e. any claim which exclusively concerns real entities) which can be proven using T_{R} together with the machinery of nominalization, could have been proved without relying upon such devices. So the machinery of nominalization is provably reliable for reasoning about what is real. Specifically, it never leads us from a true real-restricted claim to a false real-restricted claim.

---

weaker biconditional, $(\underline{v_1}, \ldots, \underline{v_n}\ \varepsilon_n\ \underline{u}) \leftrightarrow u(v_1, \ldots, v_n)$. In that case, rather than defining instantiation explicitly, they should take each instantation-predicate as a primitive, governed by schemes: $\underline{s^{\mathsf{t}}} =_{\mathsf{e}} app_n(x^{\mathsf{e}}, y_1^{\mathsf{e}}, \ldots, y_n^{\mathsf{e}}) \rightarrow ((y_1, \ldots, y_n\ \varepsilon_n\ x) \leftrightarrow s)$ and $(y_1, \ldots, y_n\ \varepsilon_n\ x) \rightarrow \exists s^{\mathsf{t}}\ \underline{s} =_{\mathsf{e}} app_n(x, y_1, \ldots, y_n)$.

[43] See Field (1980/2016, 1989/1991). In particular, we share Field's (1980/2016: P4) response to the distinction between *hermeneutic* and *revolutionary* fictionalism (see Burgess and Rosen 1997: pt. I ch. A; Stanley 2001; Burgess 2004). Our (limited) interest in that distinction can be summed up as follows: we recommend that Fregean realists should not (falsely) believe that universals exist, and so *should* engage in pretence insofar as they want to continue to use universals-discourse.

This answer also allows us to continue to speak as if universals exist, even though we deny that they do. T_{U} is a convenient tool for drawing inferences between true, real-restricted claims; and it is provably reliable, in this regard. So there is no need to jettison this tool. Indeed, the Conservativeness Theorem allows Fregean realists to pretend that universals exist, with a perfectly clear conscience. And that is what we recommend.

In the next few sections, we will illustrate T_{U}'s utility. First, we should emphasise that (Field-style) fictionalism is the live option, rather than eliminativism. It is easy to eliminate the nominalizations from *some* sentences. For example, the fictional sentence 'Socrates instantiates wisdom' can easily be replaced with the real-restricted sentence 'Socrates is wise'; indeed, T_{U} proves their equivalence (see §4.3).[44] However, it is not always possible to paraphrase nominalizations away. For example, the fictional sentence 'Socrates is not identical to wisdom' is a theorem of T_{U} (by Nom-nonreal), but there is no equivalent real-restricted sentence.

6 Some simple applications

In the remainder of this paper, we will apply our fictionalism to a variety of natural language constructions. The basic idea is this: plenty of natural English constructions seem to require us to move between types; strictly-typed logics, like PFTT, struggle with this; but our fictionalism has the resources to make it easy.

6.1 On virtues

We started §3 with this example:

(7) Wisdom is a virtue

As we explained, we think that (7) is literally false, because we do not believe in the universal wisdom. However, we also think that there is a nearby literal truth about the property $wise^{\mathtt{et}}$:

(7a) $virtue^{(\mathtt{et})\mathtt{t}}(wise^{\mathtt{et}})$

It is essential that we sharply distinguish $virtue^{(\mathtt{et})\mathtt{t}}$ from the notion of virtue at play in the original (7): the English predicate 'is a virtue' is type $\mathtt{et}$, and it expresses a type $\mathtt{et}$ property that is supposed to apply to certain universals; by contrast, $virtue^{(\mathtt{et})\mathtt{t}}$ is a type $(\mathtt{et})\mathtt{t}$ property that applies to certain type $\mathtt{et}$ properties. (If you are wondering where this new type $(\mathtt{et})\mathtt{t}$ property came from, or what it could possibly have to do with *virtue*, please hold that thought; we will come back to it in §6.2.) Unlike (7),

[44] Here (and throughout) we assume that Socrates is a real object, and that *wise* is a real property, and that both '*socrates*' and '*wise*' are atomic constants. Similar points apply for other examples in §§6–7 (e.g.: *virtue*, *believes*, *plato*, $love_{\mathtt{e}}$, and $love_{\mathtt{e}(\mathtt{et})}$).

sentence (7a) makes no mention of universals, and so we are free to accept it as a literal truth.

But now that we have (7a) as a literal truth, our fiction allows us to infer the following claim about the universal wisdom:

(7b) $\underline{\textit{virtue}}^{(\mathtt{et})\mathtt{t}}(\underline{\textit{wise}}^{\mathtt{et}})$

(Indeed, in the fiction, (7a) is identical to (7b); see §4.3.) We would like to offer (7b) as our formalization of (7). In other words, our proposal is that 'is a virtue' should be formalized in PFTT as '$\underline{\textit{virtue}}^{(\mathtt{et})\mathtt{t}}$', just as 'wisdom' should be formalized as '$\underline{\textit{wise}}^{\mathtt{et}}$'. On this proposal, (7) is literally false but assertible nonetheless, since its formalization, i.e. (7b), is true in our (conservative) fiction.

6.2 Reverse-engineering fictions

In our efforts to make (7) true within our fiction, we helped ourselves to a new type $(\mathtt{et})\mathtt{t}$ property, $\textit{virtue}^{(\mathtt{et})\mathtt{t}}$. But this might seem to raise a number of questions, for example: *What does it take for a type* $\mathtt{et}$ *property, such as* $\textit{wise}^{\mathtt{et}}$*, to satisfy* $\textit{virtue}^{(\mathtt{et})\mathtt{t}}$*?* Or to put the question another way: *What exactly is* (7a) *meant to say?* And: *What justifies our proposal to formalize 'is a virtue' as '*$\underline{\textit{virtue}}^{(\mathtt{et})\mathtt{t}}$*'?* These might seem like urgent questions for our fictionalist account of universals. In fact, they can be bypassed entirely. To explain why, we need to outline two ways of thinking about our fiction.

Our discussion in §§4–5 might suggest the following picture:

The Bolt-On Picture. We start with a real theory, T, which expresses a body of literal truths without using any nominalization-devices. We then enrich T, moving to T_{U}, by bolting on NAT's nominalization-devices. This move is justified by the Conservativeness Theorem.

If the Bolt-On Picture were taken as a description of the actual, temporal process of how humans arrive at T_{U}, then the above questions about $\textit{virtue}^{(\mathtt{et})\mathtt{t}}$ would be pressing. But that is not the point of the Bolt-On Picture. The Bolt-On Picture can be used to explain why, *given* T, we can employ T_{U} and act as if universals existed. But, of course, no one actually starts out (in the temporal sense of 'starts out') with the real theory, T, pristine and fully acceptable to Fregean realists. Rather, they arrive at T after some reflection. That process of reflection is more accurately described using an alternative picture:

The Reverse-Engineering Picture. We start out with a class of assertible sentences. Some of these sentences feature nominalized predicates in referential position. Philosophical argument, however, makes us leery of universals. So: rather than accepting the sentences in our class as literally true, we grant them the status *true in the fiction of universals*. Now our job is to *reverse-engineer* this fiction, to arrive at our real theory, T.[45]

[45] This is probably *still* an oversimplification. It assumes that every sentence in our initial class of sentences will make its way into our fiction. In fact, in the course of regimenting the real/fictional distinction, we might be led to revise our view of which sentences are really assertible; indeed, we may

This alternative Picture offers us a way around the apparently pressing questions about $virtue^{(et)t}$. Our initial class of assertible sentences includes a variety of sentences that describe certain universals as virtues, including:

(7) Wisdom is a virtue

Since (7) is assertible, its formalization should be true in our fiction. To achieve this, we hypothesize that the English predicate 'is a virtue' expresses the flattening of some real type $(et)t$ property, $x^{(et)t}$, such that $x(wise)$. It is only at the end of this process of reverse-engineering that we introduce the label '$virtue^{(et)t}$' for x: it is really just a helpful mnemonic, to remind us that this is the real type $(et)t$ property we posited in order to make sentences like (7) true in our fiction.[46]

6.3 Instantiation

Now that we have explained how to use our fiction, we would like to offer two more examples of it in action. As we noted in §4.3, NAT allows us to define a notion of instantiation, and hence to formalize a claim like 'Socrates instantiates wisdom' via:

(1b) $socrates \in \underline{wise}$

This allows us to explain the validity of various natural language inference-patterns. For example, consider this intuitively valid inference:

(1) Socrates is wise
(7) Wisdom is a virtue
(8) Therefore, Socrates has a virtue

We can formalize this inference as follows:

(1a) $wise(socrates)$
(1b) $\therefore\ socrates \in \underline{wise}$
(7b) $\underline{virtue^{(et)t}}(\underline{wise})$
(8b) $\therefore\ \exists x^{e}((socrates \in x) \wedge \underline{virtue^{(et)t}}(x))$

This argument is valid within our fiction, since (1a) implies (1b) in the fiction,[47] and then (8b) follows from (1b) and (7b) by elementary inference rules.[48] Moreover, given

need to go back and forth repeatedly, until we reach a reflective equilibrium. However, this complication does not affect our point here.

[46] Our discussion of reverse-engineering is somewhat reminiscent of *easy-road mathematical fictionalism* (e.g. Balaguer 1998: §3.2; Melia 2000; Leng 2010). Easy-road fictionalists also start with a class of assertible sentences—those delivered by mathematical science—and work backwards to (what they take to be) the literal truth. However, there is also an important difference. Easy-road fictionalists deny, or at least refuse to assert, that it is possible to state (what they take to be) the literal truth directly, without going via their fiction. By contrast, the end-product of our reverse-engineering is precisely a direct statement of (what we take to be) the literal truth.

[47] Indeed, in the fiction, they are identical_{t}.

[48] Note that T_{U} is a PFTT-theory, and hence uses the natural deduction system outlined in §A.3.

that (1a) and (7b) are both true in the fiction, the argument is not just *valid* but *sound* in the fiction.

6.4 Propositions

Our focus in this paper is on the nominalization of type `et` expressions. But our fiction can smoothly nominalize other types too. Consider the following natural language sentence:

(9) Gottlob believes that arithmetic is reducible to logic

Some philosophers and linguists take the complementizer 'that' to function as a device for nominalizing sentences (type `t`).[49] They take 'believes' to express a type `e(et)` relation between Gottlob and a *reified proposition*, i.e. the nominalization of a type `t` proposition. The argument for universals that we presented in §3.1 can now be reworked as an argument for reified propositions: (9) is true; (9) cannot be true unless 'that arithmetic reduces to logic' refers to a reified proposition; therefore at least one reified proposition exists.

We do not believe in reified propositions any more than we believe in universals, and so we want to resist this argument. We distinguish two possible lines of resistance. The first was proposed in earlier work by Trueman.[50] He suggested that, rather than being a device of nominalization, the complementizer is actually semantically vacuous. According to this suggestion, then, we should formalize (9) in PFTT as:

(9a) $believes^{\mathtt{e(tt)}}(gottlob^{\mathtt{e}}, (arithmetic\text{-}is\text{-}reducible\text{-}to\text{-}logic)^{\mathtt{t}})$

If this is the right way to formalize (9), then it does not express a relation between Gottlob and a *reified* proposition (type `e`); it expresses a relation between Gottlob and a proposition proper (type `t`).

The second line of resistance is made available by our fictionalism (and should be compared with our discussion of '$virtue^{\mathtt{(et)t}}$' in §6.1). We can grant that the complementizer in (9) is a nominalization device, but then deny that (9) is literally true; it is really only true within the fiction of universals. This result can be secured in three steps. First, we take (9a) to be a literal truth. Second, we use our fiction to infer:

(9b) $\underline{believes^{\mathtt{e(tt)}}}(gottlob^{\mathtt{e}}, \underline{(arithmetic\text{-}is\text{-}reducible\text{-}to\text{-}logic)^{\mathtt{t}}})$

Third, and finally, we offer (9b) as our formalization of (9). In other words: we propose that the English predicate 'believes' should be formalized as '$\underline{believes}^{\mathtt{e(tt)}}$'.

We do not want to take a final stand on the best way to formalize (9) itself. That said, it is worth noting that the fictionalist strategy is forced upon us in at least some cases. Consider:

[49] This view is extremely widespread; see e.g. Cresswell (1973: 166–9), Parsons (1979: 132), Künne (2003), and King et al. (2014).

[50] Trueman (2018, 2020, 2021: ch.12–13). Trueman's discussions draw heavily on Prior (1971: esp. ch.2). See also Montague (1973: 18–19) and Rosefeldt (2008).

(10) Gottlob believes logicism

'Logicism' is clearly a nominalization,[51] and so when it comes to (10), we have no choice but to deploy our fiction.

Our fiction not only copes with simple belief-reports like (9), but it can also handle iterated belief-reports, such as:

(11) Bertrand believes that Gottlob believes that arithmetic is reducible to logic

We can formalize this in either of the following ways:

(11a) $\mathit{believes}^{\mathsf{e(tt)}}(\mathit{bertrand}^{\mathsf{e}}, \mathit{believes}^{\mathsf{e(tt)}}(\mathit{gottlob}^{\mathsf{e}}, (\mathit{arithmetic\text{-}is\text{-}reducible\text{-}to\text{-}logic})^{\mathsf{t}}))$
(11b) $\underline{\mathit{believes}}^{\mathsf{e(tt)}}(\mathit{bertrand}^{\mathsf{e}}, \underline{\mathit{believes}}^{\mathsf{e(tt)}}(\mathit{gottlob}^{\mathsf{e}}, \underline{(\mathit{arithmetic\text{-}is\text{-}reducible\text{-}to\text{-}logic})^{\mathsf{t}}}))$

Fregean realists can, of course, regard (11a) as literally true. Moreover, in the fiction, it entails (11b); indeed, in the fiction, they are identical$_{\mathsf{t}}$.[52]

7 Bridge-Principles

So far, we have focussed on relatively simple natural language constructions. In this section, we will discuss two more challenging kinds of case: mixed-predication (§7.1) and pseudo-self-predication (§7.3). As we will see, these cases can be handled if we augment our fiction with certain *bridge-principles*, whose status we explore in §7.2.

7.1 Mixed-predication

Natural language allows us to construct cases of *mixed-predication*, where one and the same thing is predicated of a universal and of an ordinary object. Here are some examples:[53]

(12) Plato loves Socrates and wisdom
(13) Not only are individual electrons physical, but so is electronhood itself
(14) Mary can see roses, but not the colour red
(15) The M25 is dangerous, and so is reckless driving
(16) Hillary talked about Nancy and running for president
(17) Logicism and Ludwig both frustrated Gottlob

[51] We can use cases of mixed-predication to show that 'logicism' is a name; see e.g. (17) below.

[52] By the definition of flattening, (9a) $=_{\mathsf{t}}$ (9b); so $\mathit{believes}^{\mathsf{e(tt)}}(\mathit{bertrand}, (9a)) =_{\mathsf{t}} \mathit{believes}^{\mathsf{e(tt)}}(\mathit{bertrand}, (9b))$; so $\mathit{believes}^{\mathsf{e(tt)}}(\mathit{bertrand}, (9a)) =_{\mathsf{t}} \underline{\mathit{believes}}^{\mathsf{e(tt)}}(\mathit{bertrand}, \underline{(9b)})$, hence (11a) $=_{\mathsf{t}}$ (11b). Note that we are here substituting sentences within the the scope of propositional attitudes; that is not entirely uncontroversial, but we will not attempt to settle the proper logic for hyper-intensional contexts here.

[53] Chierchia (1982: 310–3, 1984: 8–9) seems to have been the first author to raise mixed-predication as a problem for Montagovians (focussing on gerunds and infinitives; see also Chierchia and Turner 1988: 293); we adapt (15) and (16) from him. See also Parsons (1979: 130) for an example similar to (16).

(18) This book is about Lovelace and exponentiation

Now, some of these examples may strike you as puns, on a par with the following zeugma:

(19) My mother taught me how to prove dough and theorems.

If so, our aim is not to convince you otherwise. Our point is only that (12)–(18) strike us as reasonable assertions, which illustrate a general phenomenon. We will offer a way to handle this phenomenon, focussing on the case of (12). However, our discussion naturally extends to cover any cases of mixed-predication which you do *not* want to dismiss as puns.

Unfortunately, as it stands, our fiction cannot accommodate (12). To see why, let us split (12) into two parts:

(20) Plato loves Socrates
(21) Plato loves wisdom

Sentence (20) obviously does not pose any problems. We can formalize it as:

(20a) $loves^{\mathtt{e(et)}}_{\mathtt{e}}(plato, socrates)$

This makes no mention of universals, and so we can accept it as a literal truth.

Sentence (21) is a little trickier, but we can approach it in much the same way that we approached (7) in §6.1. We help ourselves to a type $\mathtt{e((et)t)}$ relation, $loves^{\mathtt{e((et)t)}}_{\mathtt{et}}$. We then take it to be literally true that Plato bears this relation to $wise^{\mathtt{et}}$:

(21a) $loves^{\mathtt{e((et)t)}}_{\mathtt{et}}(plato, wise)$

Finally, we use the fiction of universals to infer:[54]

(21b) $\underline{loves^{\mathtt{e((et)t)}}_{\mathtt{et}}}(plato, \underline{wise})$

We think this a good formalization of (21). But, even though we now have formalizations for (20) and (21), we still cannot yet formalize (12). The trouble is that, even in the fiction, $loves_{\mathtt{e}}$ is distinct from $\underline{loves_{\mathtt{et}}}$.[55]

To solve this problem, we need to augment our fiction with a principle that, in effect, extends $loves_{\mathtt{e}}$ to include pairs of objects which stand in the $\underline{loves_{\mathtt{et}}}$ relation too. More precisely, we lay down the following *bridge-principle*:

$$(real(u^{\mathtt{e}}) \wedge real(v^{\mathtt{et}})) \rightarrow loves_{\mathtt{e}}(u, \underline{v}) =_{\mathtt{t}} loves_{\mathtt{et}}(u, v)$$

Given this bridge-principle, (21a) implies:

[54] We should repeat the assumption, from footnote 44, that we are tacitly assuming that the various wffs in these examples are atomic.

[55] Indeed, in the fiction, (20a) is true but $\underline{loves_{\mathtt{et}}}(plato, socrates)$ is false. Recall, in the fiction: $\underline{loves_{\mathtt{et}}}(x^{\mathtt{e}}, y^{\mathtt{e}})$ iff $x^{\mathtt{e}}$ and $y^{\mathtt{e}}$ are nominalizations of entities between which $loves_{\mathtt{et}}$ obtains; but *socrates* is not the nominalization of any type $\mathtt{et}$ entity (by Nom-nonreal).

(21c) $loves_{\mathtt{e}}(plato, \underline{wise})$

Together with (20a), this finally implies:

(12a) $loves_{\mathtt{e}}(plato, socrates) \wedge loves_{\mathtt{e}}(plato, \underline{wise})$

which is our formalisation, within our fiction, of (12).

7.2 Licensing the use of bridge-principles

We have dealt with mixed-predication using bridge-principles. We now need to address two questions concerning the use of such principles.

The most immediate question is: what permits us to add these bridge-principles to our fiction? Roughly put, our answer is that, if we take even a modicum of care, then adding bridge-principles to our fiction will still be conservative. To make this more precise, we must define some notions. We stipulate that a *bridge-principle* is any formula of this shape:

$$\big(real(u_1^{\alpha_1}) \wedge \ldots \wedge real(u_n^{\alpha_n})\big) \longrightarrow \mathbf{A}^{\mathtt{e}(\ldots(\mathtt{et})\ldots)}(\underline{u_1}, \ldots, \underline{u_n}) =_{\mathtt{t}} \mathbf{B}^{\alpha_1(\ldots(\alpha_n \mathtt{t})\ldots)}(u_1, \ldots, u_n)$$

where $\mathbf{A}$ and $\mathbf{B}$ are T-constants, and some type α_i is not $\mathtt{e}$. (This last clause ensures that some nominalization is, in fact, being invoked.) Then we can prove, roughly, that a set of bridge-principles, Δ, can be conservatively added to T_{U}, provided that it is impossible for the bridge-principles to conflict with each other.

Of course, this notion of 'impossibility of conflict' needs to be made more precise. So let us start by thinking about some cases where conflict is possible. Conflict is clearly possible if we insist, for example, both that $loves_{\mathtt{e}}(u^{\mathtt{e}}, \underline{v^{\mathtt{et}}}) =_{\mathtt{t}} loves_{\mathtt{et}}(u, v)$ and that $loves_{\mathtt{e}}(u^{\mathtt{e}}, \underline{v^{\mathtt{et}}}) =_{\mathtt{t}} hates_{\mathtt{et}}(u, v)$. Only slightly less clearly: imagine that $T \vdash ploves^{\mathtt{et}} =_{\mathtt{et}} loves_{\mathtt{e}}(plato)$. (Recall that we tend to write the decurried expression '$loves_{\mathtt{e}}(plato, x^{\mathtt{e}})$' in place of '$(loves_{\mathtt{e}}(plato))(x^{\mathtt{e}})$'; but here we are relying on the fact that '$loves_{\mathtt{e}}(plato)$' is a type $\mathtt{et}$ expression, roughly corresponding to the English predicate 'Plato loves …'.) In that case, conflict would be possible between a bridge-principle which mentioned $loves_{\mathtt{e}}$ and one which mentioned *ploves*. Fortunately, it is not hard to lay down a condition which rules out exactly these kinds of conflicts; this is the notion of *T-friendliness* (see Definition 7 of §E). And we can then prove that any set of T-friendly bridge-principles is conservative over T. (For full details, though, the reader will have to consult §E.)

This explains what *permits* us to introduce bridge-principles. But we should also ask: what *motivates* their introduction? For example, why should we introduce a bridge-principle between $loves_{\mathtt{e}}^{\mathtt{e}(\mathtt{et})}$ and $loves_{\mathtt{et}}^{\mathtt{e}((\mathtt{et})\mathtt{t})}$? It is tempting to answer that $loves_{\mathtt{e}}$ and $loves_{\mathtt{et}}$ are two types of loving-relation. But that answer would be strictly nonsensical, given Fregean realism, since nothing that can be said of one type can be said of another.

As in §6.2, this question can be addressed by recalling that we are *reverse-engineering* our fiction. To illustrate, consider our example of mixed-predication, and its formalization:

(12) Plato loves both Socrates and wisdom
(12a) $loves_{\mathtt{e}}(plato, socrates) \wedge loves_{\mathtt{e}}(plato, \underline{wise})$

Suppose that (12) is assertible. In that case, (12a) should be true in our fiction. Now, since $loves_{\mathtt{e}}(plato, socrates)$ is assertible and makes no reference to universals, we can accept it as a literal truth (and hence also as true in the fiction). And, given the assertibility of $loves_{\mathtt{e}}(plato, \underline{wise})$, we reverse-engineer our fiction to include some 'corresponding' relation of type $\mathtt{e((et)t)}$ which holds between *plato* and *wise*. Our bridge-principles simply formalize the intuitive talk of a 'corresponding relation'. So, these bridge-principles serve as a formal link (within the fiction) between the familiar type $\mathtt{e(et)}$ relation $loves_{\mathtt{e}}$, and the reverse-engineered type $\mathtt{e((et)t)}$ relation.

7.3 Pseudo-self-predication

Now that we have a general licence to invoke bridge-principles, we can put them to further work. For example, consider this case:

(22) Plato loves love

This is much like (21), except that it is a case of *pseudo-self-predication*, where a relation is applied to its own nominalization.[56] To handle this, as in §7.1, we introduce another type of love, $loves^{\mathtt{e((e(et))t)}}_{\mathtt{e(et)}}$, and accept the following as a literal truth:

(22a) $loves_{\mathtt{e(et)}}(plato, loves_{\mathtt{e}})$

Within our fiction, (22a) is identical$_{\mathtt{t}}$ to:

(22b) $\underline{loves_{\mathtt{e(et)}}}(plato, \underline{loves_{\mathtt{e}}})$

But $\underline{loves_{\mathtt{e(et)}}}$ is distinct from $loves_{\mathtt{e}}$, and so (22b) does not apply a relation to its own nominalization. To overcome this, we lay down another bridge-principle:

$$\big(real(u^{\mathtt{e}}) \wedge real(v^{\mathtt{e(et)}})\big) \rightarrow loves_{\mathtt{e}}(u, \underline{v}) =_{\mathtt{t}} loves_{\mathtt{e(et)}}(u, v)$$

Using this, (22a) is identical$_{\mathtt{t}}$ in the fiction to:

(22c) $loves_{\mathtt{e}}(plato, \underline{loves_{\mathtt{e}}})$

And this is a genuine case of pseudo-self-predication, in which $loves_{\mathtt{e}}$ is applied to its own nominalization.

The approach generalises. Using bridge-principles, we can easily handle other cases which have been thought to pose difficulties for strictly typed theories, such as:[57]

[56] It is only *pseudo*-self-predication: Fregean realism prohibits the application of a relation to *itself*.
[57] Example (23) is modelled off Chierchia (1984: 12, 1985: 418) and Chierchia and Turner (1988: 293).

(23) Kindness is kind
(23a) $kind_{\mathsf{e}}(\underline{kind_{\mathsf{e}}})$
(24) Plato loves loving love
(24a) $loves_{\mathsf{e}}(plato, \underline{\lambda x^{\mathsf{e}}\ loves_{\mathsf{e}}(x, \underline{loves_{\mathsf{e}}})})$

We leave it to the reader to specify the suitable bridge-principles.

8 Limits of the fiction

In §§6–7, we gave a sense of the power and flexibility of our fiction. Our hope is that this will allow us to accommodate enough of the ordinary use of nominalized predicates, to allow us to continue speaking as if there were universals in everyday and scientific contexts. Whilst there is obviously no way to *prove* that our hope has been fulfilled, we think it is extremely plausible.

However, we should also admit that there are limits to our fiction. In particular, our fiction cannot handle cases that involve nominalizing the unreal entities introduced by NAT. Here is a particularly striking example. Consider these two sentences:

(25) Socrates does not instantiate himself
(26) Socrates instantiates non-self-instantiation

We can easily formalise (25) as follows, where $\mathbf{A} \notin \mathbf{A}$ abbreviates $\neg(\mathbf{A} \in \mathbf{A})$:[58]

(25a) $socrates \notin socrates$

However, there is no adequate way to formalise (26). In the fiction, we can consider the property of non-self-instantiation, $d^{\mathsf{et}} =_{\mathsf{et}} (\lambda x^{\mathsf{e}}\ x \notin x)$; so we might offer:

(26a) $socrates \in \underline{d}$

But, since d is not real, it has no nominalization,[59] so (26a) is false.

This is, then, a limitation on NAT. But, given our aims, it is no real shortcoming. As we explained in §5, the point of our fictionalism is to explain why ordinary discourse about universals does not lead from real-restricted truths to real-restricted falsehoods, without conceding that universals really exist. We doubt that (26) is any part of the *ordinary* discourse about universals. Of course, (26) is part of the *philosophical* discourse about universals. But notoriously, philosophers *have* been led from truths to falsehoods by sentences like (26), which are just one step away from Russell's Paradox, in the form 'non-self-instantiation instantiates non-self-instantiation'.

[58] By itself, NAT does not entail (25a), for reasons noted at the end of §4.3, but see §D.

[59] For reductio, suppose $\mathsf{E}\underline{d}$. By the instantiation-scheme of §4.3, we have $(\underline{d} \in \underline{d}) =_{\mathsf{t}} (\underline{\underline{d}} \in \underline{d}) =_{\mathsf{t}} d\underline{d}$. Also, $d\underline{d} =_{\mathsf{t}} (\underline{d} \notin \underline{d})$ by *Red* and $\mathsf{E}I_{\mathsf{t}}$ (see §A.3). So $(\underline{d} \in \underline{d}) =_{\mathsf{t}} (\underline{d} \notin \underline{d})$, which is a contradiction.

9 Conclusion

Fregean realists reject the suggestion that properties are objects (see §2). Nevertheless, Fregean realism is compatible with the claim that (at least some) properties correspond to certain special objects, which we call *universals*. Moreover, various ordinary natural language constructions imply that universals exist. At the same time, however, Fregean realists have good reason to deny that there really are any universals (see §3). So, Fregean realists have good reason to embrace fictionalism about universals. We have gone some way to showing how this can be done: Fregean realists can help themselves to a provably conservative fiction of universals (see §§4–5).

This fiction allows us to provide a face-value semantics for a wide variety of natural language constructions (see §6), including cases of mixed-predication and pseudo-self-predication (see §7). Inevitably, Russell's Paradox places limits on what can be achieved within the fiction, but these limits seem somewhat recherché (see §8). Indeed, it seems that wherever one might naturally want to speak as if there were universals, Fregean realists can do just that, and in good conscience, by invoking their conservative fiction.

A PFTT's deductive system

In this appendix, we outline the formal system PFTT. This retains bivalence, but remains suitable for partial functions. It is a minor modification of Farmer's (1990) system **PF**; the main difference is that PFTT allows there to be more than two propositions.

A.1 PFTT*'s grammar*

We have two basic types: $\mathtt{e}$ and $\mathtt{t}$. For any types α and β, we have the type $(\alpha\beta)$. These are our only types. Since we have no product-types, we officially handle relations by currying. However, for readability, we often use decurried expressions. (See §1 for this, and other, notational conventions.)

For each type α, we have primitive symbols as follows:[60]

- Improper symbols: λ, (,)
- Logical constants: $\mathit{False}^{\mathtt{t}}_{\mathtt{t}}$, $\neg^{\mathtt{tt}}$, $\vee^{\mathtt{t(tt)}}$, $\wedge^{\mathtt{t(tt)}}$, $\rightarrow^{\mathtt{t(tt)}}$, $=^{\alpha(\alpha\mathtt{t})}_{\alpha}$, $\Sigma^{(\alpha\mathtt{t})\mathtt{t}}_{\alpha}$, $\Pi^{(\alpha\mathtt{t})\mathtt{t}}_{\alpha}$, $\iota^{(\alpha\mathtt{t})\alpha}_{\alpha}$
- Variables: $a^{\alpha}, \ldots, z^{\alpha}$, with subscripts as necessary
- Non-logical constants: as we see fit

[60] Farmer's (1990) **PF** allows for only two $\mathtt{t}$-entities; so he adapts an equational Henkin–Andrews system (see e.g. Andrews 2002: ch.5). To allow for multiple $\mathtt{t}$-entities, we must revert from an equational system to a system more like Church's (1940) original. We have more primitives than usual, but treat $\leftrightarrow$ as a conjunction of conditionals.

We indicate the types of expressions with superscripts. The grammar is standard:[61] every constant or variable of type α is a wff$_\alpha$; if $\mathbf{x}^\alpha$ is a variable and $\mathbf{B}^\beta$ is a wff$_\beta$, then $(\lambda\mathbf{x}^\alpha\mathbf{B}^\beta)^{\alpha\beta}$ is a wff$_{\alpha\beta}$; if $\mathbf{A}^\alpha$ is a wff$_\alpha$ and $\mathbf{B}^{\alpha\beta}$ is a wff$_{\alpha\beta}$, then $(\mathbf{BA})^\alpha$ is a wff$_\beta$; nothing else is a wff.

It is convenient to introduce some definitions:[62]

$$\forall\mathbf{x}^\alpha\mathbf{A}^{\mathtt{t}} \coloneqq \Pi_\alpha(\lambda\mathbf{x}^\alpha\mathbf{A}) \qquad Uni_\alpha^{(\alpha\mathtt{t})\mathtt{t}} \coloneqq \lambda v^{\alpha\mathtt{t}}\exists x^\alpha\forall y^\alpha(v(y) \leftrightarrow x =_\alpha y)$$
$$\exists\mathbf{x}^\alpha\mathbf{A}^{\mathtt{t}} \coloneqq \Sigma_\alpha(\lambda\mathbf{x}^\alpha\mathbf{A}) \qquad \sqsubset\mathbf{A}^\alpha \coloneqq \mathbf{A} =_\alpha \mathbf{A}$$
$$\imath\mathbf{x}^\alpha\mathbf{A}^{\mathtt{t}} \coloneqq \iota_\alpha(\lambda\mathbf{x}^\alpha\mathbf{A}) \qquad \mathbf{A}^\alpha \simeq_\alpha \mathbf{B}^\alpha \coloneqq (\sqsubset\mathbf{A} \vee \sqsubset\mathbf{B}) \rightarrow \mathbf{A} =_\alpha \mathbf{B}$$

Intuitively: Uni_α indicates that a property is uniquely instantiated; $\sqsubset\mathbf{A}$ indicates that $\mathbf{A}$ exists; and $\mathbf{A} \simeq \mathbf{B}$ indicates that they are identical if either exists.

A.2 Handling empty terms

Our system allows for empty terms. Intuitively, applying any term to an empty term causes a 'crash'. Our precise implementation of this idea follows Farmer. We recursively specify two *kinds* of types:

Definition 1: e is e-kind; t is t-kind; and $\alpha\beta$ is the same kind as β.

Intuitively, 'crashing' for e-kinds will amount to undefinedness; 'crashing' for t-kinds will amount to falsity. More precisely:

(i) e-kind expressions stand for *partial* functions. If β is e-kind then $\mathbf{B}^{\alpha\beta}\mathbf{A}^\alpha$ denotes iff: $\mathbf{B}$ denotes some b, and $\mathbf{A}$ denotes some a, and $b(a)$ is defined; in that case, $\mathbf{BA}$ denotes $b(a)$.

(ii) t-kind expressions stand for *total* functions. If $\mathbf{A}$ is empty, then $\mathbf{B}^{\alpha\mathtt{t}}\mathbf{A}^\alpha$ is false.

Clause (ii) allows us to have a negative-free logic. In fact, we will want to extend the condition on falsity, to deal with the general case of $\mathbf{B}^{\alpha\beta}\mathbf{A}^\alpha$ when β is t-kind. We use the following recursive definition:

$$False_{\alpha\beta}^{\alpha\beta} \coloneqq \lambda x^\alpha False_\beta^\beta, \text{ for all t-kind } \beta$$

So, $False_{\mathtt{t}}^{\mathtt{t}}$ is a logical constant, which we will treat as a primitive, canonical, falsity; then the various $False_{\alpha\beta}$s allow us to push higher-typed entities (ultimately) to $False_{\mathtt{t}}$. To illustrate this: suppose that c^α is non-existent; then $(=_\alpha c^\alpha)$ should be a function which yields $False_{\mathtt{t}}$ for any type α input, i.e. we want that $(=_\alpha c^\alpha) =_{\alpha\mathtt{t}} False_{\alpha\mathtt{t}}$.

[61] 'wff$_\alpha$' abbreviates 'wff of type α'. Bold capital letters stand for wffs of the indicated type. Bold lowercase letters stand for variables of the indicated type.

[62] See Church (1940: 58) for $\forall$ and $\imath$; cf. Farmer (1990) for $False_{\alpha\beta}$; and our $\sqsubset$ is based on Farmer's $\downarrow$.

A.3 PFTT*'s natural deduction system*

We provide a natural deduction system for PFTT.[63] We help ourselves to some standard classical introduction and elimination rules for $False_{\mathtt{t}}$, $\neg$, $\vee$, $\wedge$, and $\rightarrow$; here are the new rules (where '$\Vdash$' indicates a permissible inference):[64]

Red: $\mathsf{E}\mathbf{A}^{\alpha} \Vdash (\lambda \mathbf{x}^{\alpha}\mathbf{B}^{\beta})\mathbf{A} \simeq_{\beta} \mathbf{B}[^{\mathbf{A}}/\!/_{\mathbf{x}}]$ where $\mathbf{B}[^{\mathbf{A}}/\!/_{\mathbf{x}}]$ is the result of replacing every instance of $\mathbf{x}$ in $\mathbf{B}$ with $\mathbf{A}$, if neither $\mathbf{x}$ nor any of $\mathbf{A}$'s free variables are bound in $\mathbf{B}$

$\simeq$E: $\mathbf{A}^{\alpha} \simeq_{\alpha} \mathbf{B}^{\alpha}, \mathbf{C}^{\mathtt{t}} \Vdash \mathbf{C}[^{\mathbf{B}}/_{\mathbf{A}}]$ where $\mathbf{C}[^{\mathbf{B}}/_{\mathbf{A}}]$ is the result of replacing any occurrence of $\mathbf{A}$ in $\mathbf{C}$ with an occurrence of $\mathbf{B}$, if that occurrence is not immediately preceded by λ

ΠE: $\Pi_{\alpha}\mathbf{B}^{\alpha\mathtt{t}}, \mathsf{E}\mathbf{A}^{\alpha} \Vdash \mathbf{B}\mathbf{A}$

ΠI: $\mathbf{B}^{\alpha\mathtt{t}}\mathbf{x}^{\alpha} \Vdash \Pi_{\alpha}\mathbf{B}$ if $\mathbf{x}$ is not free in $\mathbf{B}$ or in any open assumption

ΣE: $\Sigma_{\alpha}\mathbf{B}^{\alpha\mathtt{t}}, \mathbf{B}\mathbf{x}^{\alpha} \rightarrow \mathbf{A}^{\mathtt{t}} \Vdash \mathbf{A}$ if $\mathbf{x}$ is not free in $\mathbf{B}$, in $\mathbf{A}$, or in any open assumption

ΣI: $\mathbf{B}^{\alpha\mathtt{t}}\mathbf{A}^{\alpha} \Vdash \Sigma_{\alpha}\mathbf{B}$

ιI: $Uni_{\alpha}\mathbf{A}^{\alpha\mathtt{t}} \Vdash \mathbf{A}(\iota_{\alpha}\mathbf{A})$

$\iota\mathrm{E}_{\mathtt{e}}$: $\mathsf{E}(\iota_{\alpha}\mathbf{A}^{\alpha\mathtt{t}}) \Vdash Uni_{\alpha}\mathbf{A}$ if α is e-kind

$\iota\mathrm{E}_{\mathtt{t}}$: $(\iota_{\alpha}\mathbf{A}^{\alpha\mathtt{t}}) \neq_{\alpha} False_{\alpha} \Vdash Uni_{\alpha}\mathbf{A}$ if α is t-kind

$\mathsf{E}\mathrm{I}_{\mathtt{e}}$: $\Vdash \mathsf{E}\mathbf{A}^{\alpha}$ if $\mathbf{A}$ is a variable, a logical constant, or a λ-term

$\mathsf{E}\mathrm{I}_{\mathtt{t}}$: $\Vdash \mathsf{E}\mathbf{A}^{\alpha}$ if α is t-kind

$Cr_{\mathtt{e}}$: $\mathsf{E}(\mathbf{B}^{\alpha\beta}\mathbf{A}^{\alpha}) \Vdash \mathsf{E}\mathbf{B}$ if β is e-kind

$Cr_{\mathtt{e}}$: $\mathsf{E}(\mathbf{B}^{\alpha\beta}\mathbf{A}^{\alpha}) \Vdash \mathsf{E}\mathbf{A}$ if β is e-kind

$Cr_{\mathtt{t}}$: $\neg\mathsf{E}\mathbf{A}^{\alpha} \Vdash \mathbf{B}^{\alpha\beta}\mathbf{A}^{\alpha} =_{\beta} False_{\beta}$ if β is t-kind

Ext: $(\mathbf{A}^{\alpha\beta}\mathbf{x}^{\alpha}) \simeq_{\beta} (\mathbf{B}^{\alpha\beta}\mathbf{x}) \Vdash \mathbf{A} \simeq_{\alpha\beta} \mathbf{B}$ if $\mathbf{x}$ is not free in $\mathbf{A}$, in $\mathbf{B}$, or in any open assumption

To illustrate the character of this deductive system, here is a useful theorem scheme, which provides a generalization of higher-order Cantor's Theorem.[65] Our informal proof is easily formalizable within the strict deductive system we just outlined:

Theorem 2 (for any α, β)**:** Fix a function $f^{(\alpha\beta)\alpha}$, a relation $r^{\alpha(\alpha\mathtt{t})}$, and distinct type β entities $0^{\beta} \neq_{\beta} 1^{\beta}$. Consider the function $c^{\alpha\beta}$ such that:[66]

$$cx^{\alpha} := \begin{cases} 1 & \text{if } \exists v^{\alpha\beta}(r(x, fv) \wedge vx =_{\beta} 0) \\ 0 & \text{otherwise} \end{cases}$$

If $r(fc, fc)$, then $c(fc) =_{\beta} 1$, so that also $\exists d^{\alpha\beta}(r(fc, fd) \wedge d(fc) =_{\beta} 0)$.

Proof. For reductio, suppose $c(fc) =_{\beta} 0$; so $\forall v^{\alpha\beta}(r(fc, fv) \rightarrow v(fc) \neq_{\beta} 0)$. Instantiating, $r(fc, fc) \rightarrow c(fc) \neq_{\beta} 0$. By assumption, $r(fc, fc)$; so $c(fc) \neq_{\beta} 0$, a contradiction. Discharging the reductio, $c(fc) =_{\beta} 1$. □

[63] Farmer (1990) offers a Hilbert-style axiomatization.

[64] So we write e.g. $\mathbf{A}, \mathbf{B} \Vdash \mathbf{C}$ for what you might write in a Gentzen-style system as $\frac{\mathbf{A} \quad \mathbf{B}}{\mathbf{C}}$. Since $False_{\mathtt{t}}$ is our canonical absurdity, the rule *ex falso* might be given as: $False_{\mathtt{t}} \Vdash$.

[65] Compare e.g. Linnebo (forthcoming: §3).

[66] i.e. $c =_{\alpha\beta} \lambda x^{\alpha} \iota z^{\beta}((\phi(x) \rightarrow z =_{\beta} 1) \wedge (\neg\phi(x) \rightarrow z =_{\beta} 0))$, with $\phi(x)$ abbreviating $\exists v(r(x, fv) \wedge vx =_{\beta} 0)$.

Corollary 3: If $f^{(\alpha\mathtt{e})\alpha}$ is total, then f is not injective. Indeed, in this case there are non-coextensive $c^{\alpha\mathtt{e}}$ and $d^{\alpha\mathtt{e}}$ such that $fc =_\alpha fd$.

Proof. Use Theorem 2, with β as $\mathtt{t}$, $0^{\mathtt{t}}$ as $False_{\mathtt{t}}$, $1^{\mathtt{t}}$ as $\neg False_{\mathtt{t}}$, and $r^{\alpha(\alpha\mathtt{t})}$ as $=_\alpha$. Since $f^{(\alpha\mathtt{t})\alpha}$ is total, $fc =_\alpha fc$. So $c(fc) =_{\mathtt{t}} \neg False_{\mathtt{t}}$, and there is $d^{\mathtt{et}}$ such that $fc =_\alpha fd$ and $d(fc) =_{\mathtt{t}} False_{\mathtt{t}}$. Now $\neg(c(fc) \leftrightarrow d(fc))$. □

This corollary is the unsurprising version of Cantor's Theorem mentioned in §3.4.

A.4 Henkin semantics for PFTT

PFTT has a (sound and complete) Henkin semantics. Again, this follows Farmer closely. Note that the Henkin semantics assigns wffs, of all types, to *objects* from set theory with urelements. So we do *not* regard the Henkin semantics as our intended semantics for PFTT (see §2.1, especially footnote 6); rather, we regard it as a useful mathematical instrument.

A PFTT-*interpretation*, $\mathcal{M}$, comprises non-empty domains M_α for each type α, with a particular subset $M_{\mathrm{des}} \subseteq M_{\mathtt{t}}$, and a particular entity F. Intuitively, M_{des} will be our 'designated values' (our true propositions) and F will be a canonical falsehood. We let $M = \bigcup_\alpha M_\alpha$, and insist:

(1) $\mathrm{F} \in M_{\mathtt{t}} \setminus M_{\mathrm{des}}$
(2) If α is $\mathtt{e}$-kind, then $M_{\alpha\beta}$ is a set of partial functions from M_α to M_β
(3) If α is $\mathtt{t}$-kind, then $M_{\alpha\beta}$ is a set of total functions from M_α to M_β
(4) If α is $\mathtt{t}$-kind, then $\mathrm{F}_\alpha^{\mathcal{M}} \in M_\alpha$, where $\mathrm{F}_{\mathtt{t}}^{\mathcal{M}} := \mathrm{F}$ and $\mathrm{F}_{\alpha\beta}^{\mathcal{M}}(a) := \mathrm{F}_\beta^{\mathcal{M}}$ for all $a \in M_\alpha$
(5) If a constant $\mathbf{C}^\alpha$ is assigned, it is assigned to some $\mathbf{C}^{\mathcal{M}} \in M_\alpha$; and $\mathbf{C}^\alpha$ must be assigned if α is $\mathtt{t}$-kind
(6) Every logical constant is assigned; their assignments meet these rules:[67]

$$
\begin{aligned}
False_{\mathtt{t}}^{\mathcal{M}} &= \mathrm{F} = \mathrm{F}_{\mathtt{t}}^{\mathcal{M}} \\
\neg^{\mathcal{M}}(a) &\in M_{\mathrm{des}} \text{ iff } a \notin M_{\mathrm{des}} \\
(a \vee^{\mathcal{M}} b) &\in M_{\mathrm{des}} \text{ iff either } a \in M_{\mathrm{des}} \text{ or } b \in M_{\mathrm{des}} \\
(a \wedge^{\mathcal{M}} b) &\in M_{\mathrm{des}} \text{ iff both } a \in M_{\mathrm{des}} \text{ and } b \in M_{\mathrm{des}} \\
(a \rightarrow^{\mathcal{M}} b) &\in M_{\mathrm{des}} \text{ iff either } a \notin M_{\mathrm{des}} \text{ or } b \in M_{\mathrm{des}} \\
(a =_\alpha^{\mathcal{M}} b) &\in M_{\mathrm{des}} \text{ iff } a = b \\
\Sigma_\alpha^{\mathcal{M}}(b) &\in M_{\mathrm{des}} \text{ iff } b(x) \in M_{\mathrm{des}} \text{ for some } x \in M_\alpha \\
\Pi_\alpha^{\mathcal{M}}(b) &\in M_{\mathrm{des}} \text{ iff } b(x) \in M_{\mathrm{des}} \text{ for all } x \in M_\alpha \\
\iota_\alpha^{\mathcal{M}}(b) &= \text{the } x \in M_\alpha \text{ such that } b(x) \in M_{\mathrm{des}}, \text{ if there is one} \\
\iota_\alpha^{\mathcal{M}}(b) &\text{ crashes, if there is not}
\end{aligned}
$$

[67] We assume suitable domains. That is: the domain of $\neg^{\mathcal{M}}$ is $M_{\mathtt{t}}$; the domain of $(\Pi_\alpha)^{\mathcal{M}}$ and $(\iota_\alpha)^{\mathcal{M}}$ is $M_{(\alpha\mathtt{t})}$; etc. We tend to leave this assumption implicit in what follows.

Here, and in what follows, the sense of 'crashes' is as in §A.2: so $\iota_\alpha^{\mathcal{M}}(b)$ is *undefined* if α is e-kind; and $\iota_\alpha^{\mathcal{M}}(b) = \text{F}_\alpha^{\mathcal{M}}$ if α is t-kind.

Given a PFTT-interpretation $\mathcal{M}$, a *variable assignment* for $\mathcal{M}$ maps every type α variable to an element of M_α. Where σ is a variable assignment for $\mathcal{M}$, this is extended to a (partial)[68] function providing values for $\mathbf{A}_\sigma^{\mathcal{M}}$, for all wffs $\mathbf{A}$, via these recursive clauses. (We omit the superscript '$\mathcal{M}$' where it is obvious from context.) The clauses for variables, $\mathbf{x}^\alpha$, and constants, $\mathbf{C}^\alpha$, are obvious:

$\mathbf{x}_\sigma = \sigma(\mathbf{x})$	
$\mathbf{C}_\sigma = \mathbf{C}^{\mathcal{M}}$	if $\mathbf{C}$ is assigned in $\mathcal{M}$
$\mathbf{C}_\sigma$ is undefined	otherwise

Where possible, we distribute assignments over application. So, with $(\mathbf{B}^{\alpha\beta}\mathbf{A}^\alpha)$:

$(\mathbf{BA})_\sigma = \mathbf{B}_\sigma\mathbf{A}_\sigma$	if $\mathbf{A}_\sigma$, $\mathbf{B}_\sigma$ and $\mathbf{B}_\sigma\mathbf{A}_\sigma$ are all defined
$(\mathbf{BA})_\sigma$ crashes	otherwise

where, recall, crashing amounts to: if β is e-kind then $(\mathbf{BA})_\sigma$ is undefined; if β is t-kind then $(\mathbf{BA})_\sigma = \text{F}_\beta^{\mathcal{M}}$. (Note: when β is t-kind, clause (4) of our semantics guarantees that both $\mathbf{B}_\sigma$ and $\mathbf{B}_\sigma\mathbf{A}_\sigma$ are defined.) Finally, we consider λ-terms: $(\lambda\mathbf{x}^\alpha\mathbf{B}^\beta)_\sigma$ is the function in $M_{\alpha\beta}$ such that for all $a \in M_a$:

$(\lambda\mathbf{xB})_\sigma(a) = \mathbf{B}_{\sigma[\mathbf{x}:a]}$	if $\mathbf{B}_{\sigma[\mathbf{x}:a]}$ is defined
$(\lambda\mathbf{xB})_\sigma(a)$ is undefined	otherwise

where $\sigma[\mathbf{x}{:}a]$ is the variable assignment which differs from σ, if at all, by mapping $\mathbf{x}$ to a.

A PFTT-interpretation $\mathcal{M}$ is a PFTT-*structure* iff, for every variable assignment σ and every $\mathbf{A}^\alpha$: if α is e-kind then either $\mathbf{A}_\sigma^\alpha$ is undefined or $\mathbf{A}_\sigma^\alpha \in M_\alpha$; and if α is t-kind then $\mathbf{A}_\sigma^\alpha \in M_\alpha$.

When $\mathcal{M}$ is a PFTT-structure and $\mathbf{A}$ is a closed wff, $\mathbf{A}_\sigma^{\mathcal{M}}$ does not depend on the choice of σ; so we can write simply $\mathbf{A}^{\mathcal{M}}$. We write $\mathcal{M} \vDash \mathbf{A}^\text{t}$ iff $\mathbf{A}^{\mathcal{M}} \in M_{\text{des}}$. We write $\mathcal{M} \vDash T$ iff $\mathcal{M} \vDash \mathbf{A}^\text{t}$ for all $\mathbf{A}^\text{t} \in T$; in this case, we can say that $\mathcal{M}$ is a model of T.

The natural deduction system is provably sound and complete for PFTT-structures. This can be shown by making minor adjustments to Farmer's proof of soundness and completeness for his system **PF**.

A.5 Fineness of grain in PFTT

To repeat: PFTT is a minor adjustment to Farmer's **PF**, which modifies Church's theory of types. Our main departure is to allow more than two t entities. We should briefly comment on our reasons for this departure.

[68] Farmer (1990) provides a total function by having a 'default' value of $\perp \notin M$ such that, where we would say that $\mathbf{A}_\sigma^{\mathcal{M}}$ is undefined, he lets $\mathbf{A}_\sigma^{\mathcal{M}} = \perp$.

One might worry that PFTT's rule *Ext* makes the framework rather coarse-grained. For example, the apparently analogous principle of second-order logic, $\forall x(Fx \leftrightarrow Gx) \rightarrow F = G$, identifies all coextensive properties. However, since PFTT allows for more than two $\mathtt{t}$ entities, its rule *Ext* does not have the same consequence. To illustrate, in PFTT: $cordate^{\mathtt{et}}$ and $renate^{\mathtt{et}}$ are distinct properties iff there is some $x^{\mathtt{e}}$ such that $cordate(x) \neq_{\mathtt{t}} renate(x)$; since PFTT allows that there can be more than two propositions, this situation is compatible with $\forall x^{\mathtt{e}}(cordate(x) \leftrightarrow renate(x))$.

Indeed, PFTT deliberately says as little as possible about what it takes for propositions to be identical. If you want to say more about how finely grained you think propositions should be, then you are free to do so. For example: perhaps you favour *Booleanism*, according to which $u^{\mathtt{t}} =_{\mathtt{t}} (\neg\neg u)$, and $(\neg u^{\mathtt{t}} \wedge \neg v^{\mathtt{t}}) =_{\mathtt{t}} \neg(u \vee v)$, and so forth.[69] Feel free to add such claims; in the interests of inclusivity, we will remain silent about them.[70] Our desire for inclusivity also explains why PFTT has so many distinct logical primitives.[71]

B The fictional theory

As explained in §5, the Fregean realist's fictional theory, T_{U}, is obtained by adding NAT to the result of explicitly restricting the Fregean realist's theory, T, to real entities. Having defined NAT in §4.2, it only remains to define a recursive translation, R, such that $\mathbf{A}_{\text{R}}$ can be thought of as '$\mathbf{A}$ as restricted to real entities'. We define the recursive translation in this appendix, justifying the clauses in small groups. We start with the easiest:

$\mathbf{C}_{\text{R}}$ is $\mathbf{C}$, if $\mathbf{C}$ is a variable or a constant other than Σ, Π, ι or $=$

We do not really have any choice concerning variables or non-logical constants. For the sentential connectives, i.e. $False_{\mathtt{t}}$, $\neg$, $\wedge$, $\vee$, $\rightarrow$, we could have provided a more complicated definition. However, by NAT's scheme Prop-real, no new (non-real) propositions are added when we move to the fictional theory; so they should just be translated verbatim. We next consider the quantifier-like constants:

[69] See Dorr (2016: 62–70) for discussion.

[70] Note, though: if you want these new principles also to hold in the fiction, T_{U}, you may need to adjust the proof of the Conservativeness Theorem (albeit in routine ways).

[71] That said, we should flag an area where we coarse-grain more than some would like. Roughly: by $Cr_{\mathtt{t}}$, 'Freya is divine' and 'Odin is ticklish' both express the same proposition, $False_{\mathtt{t}}$. Greater fine-graining for such crashes would require making substantial adjustments to PFTT.

$$(\Sigma_\alpha)_{\mathrm{R}} \text{ is } \lambda v^{\alpha\mathtt{t}}\imath x^{\mathtt{t}}\big((real(v) \rightarrow x =_{\mathtt{t}} (\exists z^{\alpha}(real(z) \wedge vz))) \wedge (\neg real(v) \rightarrow x =_{\mathtt{t}} False_{\mathtt{t}})\big)$$

$$(\Pi_\alpha)_{\mathrm{R}} \text{ is } \lambda v^{\alpha\mathtt{t}}\imath x^{\mathtt{t}}\big((real(v) \rightarrow x =_{\mathtt{t}} (\forall z^{\alpha}(real(z) \rightarrow vz))) \wedge (\neg real(v) \rightarrow x =_{\mathtt{t}} False_{\mathtt{t}})\big)$$

$$(\iota_\alpha)_{\mathrm{R}} \text{ is } \lambda v^{\alpha\mathtt{t}}\imath x^{\alpha}\big((real(v) \rightarrow x =_{\alpha} \imath y^{\alpha}(real(y) \wedge vy)) \wedge (\neg real(v) \rightarrow x =_{\alpha} False_{\alpha})\big), \text{ if } \alpha \text{ is } \mathtt{t}\text{-kind}$$

$$(\iota_\alpha)_{\mathrm{R}} \text{ is } \lambda v^{\alpha\mathtt{t}}\imath x^{\alpha}\big((real(v) \rightarrow x =_{\alpha} \imath y^{\alpha}(real(y) \wedge vy)) \wedge (\neg real(v) \rightarrow x =_{\alpha} \imath z^{\alpha} False_{\mathtt{t}})\big), \text{ if } \alpha \text{ is } \mathtt{e}\text{-kind}$$

The intuitive idea behind these definitions is that a real-restricted quantifier should have its domain restricted to the real entities. In the particular case of existential quantification, this comes down to two ideas: (i) if $v^{\alpha\mathtt{t}}$ is a real property, then "something is v" should be really true iff some *real* thing is v; (ii) if $v^{\alpha\mathtt{t}}$ is a non-real property, then "something is v" should not really be true. Since $\Sigma_\alpha v^{\alpha\mathtt{t}}$ can be rewritten as $\exists x^{\alpha} v^{\alpha\mathtt{t}} x$, ideas (i) and (ii) respectively motivate the two conjuncts in the definition of $(\Sigma_\alpha)_{\mathrm{R}}$. A similar line of thought justifies our clauses Π and ι. Next, we consider = and λ-terms:

$$(=_\alpha)_{\mathrm{R}} \text{ is } \imath v^{\alpha(\alpha\mathtt{t})}\big(real(v) \wedge \forall x^{\alpha}\forall y^{\alpha}((real(x) \wedge real(y)) \rightarrow v(x,y) =_{\mathtt{t}} (x =_{\alpha} y))\big)$$

$$(\lambda \mathbf{x}^{\alpha}\mathbf{B}^{\beta})_{\mathrm{R}}^{\alpha\beta} \text{ is } \imath v^{\alpha\beta}\big(real(v) \wedge \forall y^{\alpha}(real(y) \rightarrow vy \simeq_{\beta} (\lambda \mathbf{x}\mathbf{B}_{\mathrm{R}})y)\big)$$

The intuitive idea is that the real-restriction of fictional-= should be the (unique) real entity which agrees precisely with fictional-= over all real entities; likewise for λ-terms. Finally, R distributes over application:

$$(\mathbf{B}^{\alpha\beta}\mathbf{A}^{\alpha})_{\mathrm{R}} \text{ is } (\mathbf{B}_{\mathrm{R}}\mathbf{A}_{\mathrm{R}})$$

This completes our definition of R, and it allows us to define T_{U} from T:

Definition 4: Let T be any PFTT theory in some signature, $\mathscr{L}$, which is disjoint from NAT's non-logical vocabulary. T's *fictionalization* is then $T_{\mathrm{U}} = \text{NAT} \cup \{\mathbf{A}_{\mathrm{R}}^{\mathtt{t}} : T \vdash \mathbf{A}\}$.

C Conservativeness

We can now state and prove our main result:

Conservativeness Theorem: Let T be a PFTT theory in some signature, $\mathscr{L}$, which is disjoint from NAT's non-logical vocabulary. If $T_{\mathrm{U}} \vdash \mathbf{A}_{\mathrm{R}}$ then $T \vdash \mathbf{A}$, for any $\mathscr{L}$-sentence $\mathbf{A}^{\mathtt{t}}$.

We prove this result via *expansion-conservation*.[72] We will show how to transform any Henkin model, $\mathcal{M}$, into a richer model, $\mathcal{M}^*$, so that the following holds:

[72] See e.g. Button and Walsh (2018: 60–2).

Lemma 5: For any $\mathcal{M}$ with signature $\mathscr{L}$:

(1) $\mathcal{M}^* \vDash \mathbf{A}_{\text{R}}$ iff $\mathcal{M} \vDash \mathbf{A}$, for any $\mathscr{L}$-sentence $\mathbf{A}^{\mathtt{t}}$; and
(2) $\mathcal{M}^*$ satisfies NAT.

The Conservativeness Theorem will then follow straightforwardly:

Proof of Conservativeness Theorem from Lemma 5. Suppose that $T_{\text{U}} \vdash \mathbf{A}_{\text{R}}$. By Soundness, $T_{\text{U}} \vDash \mathbf{A}_{\text{R}}$. Let $\mathcal{M}$ be any PFTT-structure such that $\mathcal{M} \vDash T$; then $\mathcal{M}^* \vDash T_{\text{U}}$ by Lemma 5 and hence $\mathcal{M}^* \vDash \mathbf{A}_{\text{R}}$, so that $\mathcal{M} \vDash \mathbf{A}$ by Lemma 5. Generalising, $T \vDash \mathbf{A}$. By Completeness, $T \vdash \mathbf{A}$. □

It just remains to explain how to construct $\mathcal{M}^*$ from $\mathcal{M}$, and to prove Lemma 5. The basic idea is simple. For each higher-type entity in $\mathcal{M}$, we create a new object to serve as its nominalization; we close under all possible (partial) functions; then we interpret the new vocabulary in the most obvious way. Admittedly, the details are fiddly, and spelling them out will take several steps (and the rest of this appendix). But there is nothing essentially more complicated than this very simple idea.

Proof of Lemma 5. In what follows, $\mathcal{M}$ can be any PFTT-structure in some signature $\mathscr{L}$. Recall that M_α is the set of entities which are values of type α variables in $\mathcal{M}$, with $M = \bigcup_\alpha M_\alpha$. We describe the construction of $\mathcal{M}^*$ from $\mathcal{M}$ in several steps. To avoid a rash of asterisks, we will refer to $\mathcal{M}^*$ as $\mathcal{N}$, but this should not obscure that $\mathcal{N}$ functionally depends on $\mathcal{M}$. Our proof has five steps.

Step 1. Denizens of $\mathcal{N}$. We will want $\mathcal{N}$ to have the same propositions as $\mathcal{M}$, and to treat the same propositions as designated. So we stipulate:

$$N_{\mathtt{t}} \coloneqq M_{\mathtt{t}}$$
$$N_{\text{des}} \coloneqq M_{\text{des}}$$

However, we will want $\mathcal{N}$ to contain some new nonreal objects, i.e. our universals. To this end, for each $\alpha \neq \mathtt{e}$, we fix simultaneously a set U_α and a bijection $\mu_\alpha : M_\alpha \longrightarrow U_\alpha$. Intuitively, U_α will supply the universals obtained from nominalizing type α entities. We also insist that M and all the U_αs are pairwise disjoint, and (for convenience) that $\mu_{\mathtt{e}}$ is the identity function on $M_{\mathtt{e}}$. Our type $\mathtt{e}$ domain is then:

$$N_{\mathtt{e}} \coloneqq M_{\mathtt{e}} \cup \bigcup_{\alpha \neq \mathtt{e}} U_\alpha$$

We then flesh out the domains of complex types as richly as possible:

$$N_{\alpha\beta} \coloneqq \begin{cases} \text{the set of all partial functions from } N_\alpha \text{ to } N_\beta & \text{if } \beta \text{ is } \mathtt{e}\text{-kind} \\ \text{the set of all total functions from } N_\alpha \text{ to } N_\beta & \text{if } \beta \text{ is } \mathtt{t}\text{-kind} \end{cases}$$

Step 2. Picking out 'real' entities. We now isolate some of $\mathcal{N}$'s denizens as 'real'. Roughly, these should just be the denizens of M_α. However, that idea is too rough.

Where β is $\mathtt{t}$-kind, a function of type $\alpha\beta$ must be total; but the functions of $M_{\alpha\beta}$ will not be total for N_α, since N_α will contain entities not in M_α. We must make these functions total.

To achieve this, we provide a recursive construction, simultaneously defining sets, R_α, for each type α and a function, $\star$. Intuitively, R_α comprises the 'real' entities of type α, and $\star$ provides a bijection $M_\alpha \longrightarrow R_\alpha$; this allows us to treat $a_\star$ as an ersatz for $a \in M$.

We define $R_\mathtt{e} \coloneqq M_\mathtt{e}$ and $R_\mathtt{t} \coloneqq M_\mathtt{t}$. Then, for each $a \in M_\mathtt{e} \cup M_\mathtt{t}$, we stipulate that $a_\star \coloneqq a$. For complex types: having defined R_α, R_β, and $\star$ over the types α and β, we will define $a_\star$ for each $a \in M_{\alpha\beta}$ as a function with domain N_α, stipulating:

$$\begin{array}{rl} a_\star(x_\star) \coloneqq (a(x))_\star & \text{if } a(x) \text{ is defined} \\ a_\star(y) \text{ crashes} & \text{otherwise} \end{array}$$

As before (and throughout), to say that $a_\star(y)$ crashes is to say: $a_\star(y)$ is undefined if β is $\mathtt{e}$-kind; and $a_\star(y) = \mathrm{F}^{\mathcal{N}}_\beta$ if β is $\mathtt{t}$-kind. Finally, we define:

$$R_{\alpha\beta} \coloneqq \{a_\star : a \in M_{\alpha\beta}\}$$

A routine induction confirms that this is well-defined and that, for each α, restricting $\star$ to M_α is a bijection $M_\alpha \longrightarrow R_\alpha$.

Step 3. Non-logical constants. When $\mathbf{C}$ is a nonlogical $\mathscr{L}$-constant, we stipulate

$$\begin{array}{rl} \mathbf{C}^{\mathcal{N}} \coloneqq (\mathbf{C}^{\mathcal{M}})_\star & \text{if } \mathbf{C} \text{ is assigned in } \mathcal{M} \\ \mathbf{C}^{\mathcal{N}} \text{ crashes} & \text{otherwise} \end{array}$$

We now turn to NAT's non-logical constants. For each α, we stipulate:

$$\begin{array}{rl} \mathit{real}^{\mathcal{N}}_\alpha(a_\star) \coloneqq (a =^{\mathcal{M}}_\alpha a) & \text{if } a_\star \in R_\alpha \\ \mathit{real}^{\mathcal{N}}_\alpha(b) \coloneqq \mathrm{F} & \text{if } b \in N_\alpha \setminus R_\alpha \end{array}$$

We handle nominalization by stipulating, for each $\alpha \neq \mathtt{e}$:

$$\begin{array}{rl} \mathit{nom}^{\mathcal{N}}_\alpha(a_\star) \coloneqq \mu_\alpha(a) & \text{if } a_\star \in R_\alpha \\ \mathit{nom}^{\mathcal{N}}_\alpha(b) \text{ crashes} & \text{if } b \in N_\alpha \setminus R_\alpha \end{array}$$

For readability, in what follows, we write $\underline{a}$ for $\mathit{nom}^{\mathcal{N}}_\alpha(a)$ if $a \in N_\alpha$ with $\alpha \neq \mathtt{e}$, and also write $\underline{a}$ as an alternative to a when $a \in N_\mathtt{e}$. To handle application, we stipulate that, for all $\alpha_1, \ldots, \alpha_n, \beta$:[73]

[73] Since app_n is $\mathtt{e}$-kind, this leaves a tiny amount of choice over implementation. Specifically, suppose that $\mathit{app}^{\mathcal{N}}_2(d, x)$ should crash for all x: then we can *either* let $\mathit{app}^{\mathcal{N}}_2(d)$ be the trivial partial function which crashes for every input, *or* let $\mathit{app}^{\mathcal{N}}_2(d)$ crash. For concreteness, choose the former.

$$app_n^{\mathcal{N}}(\underline{b}, \underline{a_1}, \ldots, \underline{a_n}) := \underline{b(a_1, \ldots, a_n)} \quad \text{if } b \in N_{\alpha_1(\ldots(\alpha_n\beta)\ldots)}, \text{ and each } a_i \in N_{\alpha_i}, \text{ and } \underline{b} \text{ and each } \underline{a_i} \text{ exist, and } b(a_1, \ldots, a_n) \text{ is defined}$$

$$app_n^{\mathcal{N}}(d, c_1, \ldots, c_n) \text{ crashes} \quad \text{in all other cases}$$

Step 4. Logical constants. It remains to interpret the logical constants. Since $N_{\mathsf{t}} = M_{\mathsf{t}}$, we retain the interpretations of sentential connectives, i.e.:

$$False_{\mathsf{t}}^{\mathcal{N}} := False_{\mathsf{t}}^{\mathcal{M}} \quad \neg^{\mathcal{N}} := \neg^{\mathcal{M}} \quad \vee^{\mathcal{N}} := \vee^{\mathcal{M}} \quad \wedge^{\mathcal{N}} := \wedge^{\mathcal{M}} \quad \rightarrow^{\mathcal{N}} := \rightarrow^{\mathcal{M}}$$

Concerning $=$, for each α, we stipulate that:

$$\begin{aligned}
(a_\star =_\alpha^{\mathcal{N}} b_\star) &:= (a =_\alpha^{\mathcal{M}} b) && \text{if } a_\star, b_\star \in R_\alpha \\
(\underline{a_\star} =_{\mathsf{e}}^{\mathcal{N}} \underline{b_\star}) &:= (a =_\alpha^{\mathcal{M}} b) && \text{if } a_\star, b_\star \in R_\alpha \text{ and } \alpha \neq \mathsf{e} \\
(c =_\alpha^{\mathcal{N}} c) &:= \neg^{\mathcal{M}}(\mathrm{F}) && \text{if } c \in N_\alpha \setminus R_\alpha \text{ and } \alpha \neq \mathsf{e} \\
(c =_\alpha^{\mathcal{N}} d) &:= \mathrm{F} && \text{in all other cases}
\end{aligned}$$

We must handle Σ and Π more carefully (since we want to guarantee *Fact 2*, below). For each $a \in M_{\alpha\mathsf{t}}$, let a_Σ be the function which arises by "restricting" $a_\star$ to real inputs, i.e., let $a_\Sigma(x) := (real_\alpha^{\mathcal{N}}(x) \wedge^{\mathcal{N}} a_\star(x))$ for each $x \in N_\alpha$. Now stipulate:

$$\begin{aligned}
\Sigma_\alpha^{\mathcal{N}}(a_\Sigma) &:= \Sigma_\alpha^{\mathcal{M}}(a) && \text{if } a \in M_{\alpha\mathsf{t}} \\
\Sigma_\alpha^{\mathcal{N}}(b) &:= \neg^{\mathcal{M}}(\mathrm{F}) && \text{if } b(x) \in N_{\mathrm{des}} \text{ for some } x \in M_\alpha, \text{ and } b \neq a_\Sigma \text{ for any } a \in M_{\alpha\mathsf{t}} \\
\Sigma_\alpha^{\mathcal{N}}(b) &:= \mathrm{F} && \text{in all other cases}
\end{aligned}$$

Similarly, let $a_\Pi(x) := (real_\alpha^{\mathcal{N}}(x) \rightarrow^{\mathcal{N}} a_\star(x))$ for each $x \in N_\alpha$, and stipulate:

$$\begin{aligned}
\Pi_\alpha^{\mathcal{N}}(a_\Pi) &:= \Pi_\alpha^{\mathcal{M}}(a) && \text{if } a \in M_{\alpha\mathsf{t}} \\
\Pi_\alpha^{\mathcal{N}}(b) &:= \neg^{\mathcal{M}}(\mathrm{F}) && \text{if } b(x) \in N_{\mathrm{des}} \text{ for all } x \in M_\alpha, \text{ and } b \neq a_\Pi \text{ for any } a \in M_{\alpha\mathsf{t}} \\
\Pi_\alpha^{\mathcal{N}}(b) &:= \mathrm{F} && \text{in all other cases}
\end{aligned}$$

Finally, define each $\iota_\alpha^{\mathcal{N}}$ exactly as instructed by clause (6) of §A.4. This completes the construction of $\mathcal{N}$. It is easy to check that $\mathcal{N}$ is a PFTT-structure, and it is *standard* by construction.

Step 5. Confirming Lemma 5. By construction, $\mathcal{N} \vDash$ NAT. This delivers Lemma 5.2. To secure Lemma 5.1, we first establish two facts:

Fact 1. If $a, b \in R_{\alpha\beta}$ and $a(x) \simeq b(x)$ for all $x \in R_\alpha$,[74] then $a = b$.
Fact 2. $(\mathbf{C}^{\mathcal{M}})_\star \simeq (\mathbf{C}_{\mathrm{R}})^{\mathcal{N}}$ for each T-constant $\mathbf{C}$, whether logical or non-logical.

[74] i.e. either $a(x) = b(x)$, or both $a(x)$ and $b(x)$ are undefined.

We constructed $\mathcal{N}$ to satisfy Fact 1; the non-logical cases of Fact 2 are given explicitly, and the logical cases can be easily check, using Fact 1 in the case of identity.

Using these Facts, we can secure Lemma 5.1 as follows. Where σ is any variable assignment on $\mathcal{M}$, let $\sigma_\star$ be the assignment on $\mathcal{N}$ given by $\sigma_\star(x) = (\sigma(x))_\star$. Relying on Facts 1 and 2, a routine induction shows that $(\mathbf{A}^{\mathcal{M}}_{\sigma})_\star \simeq (\mathbf{A}_{\mathrm{R}})^{\mathcal{N}}_{\sigma_\star}$ for each $\mathscr{L}$-wff $\mathbf{A}^\alpha$. So, in particular, $\mathcal{M} \vDash \mathbf{A}^{\mathsf{t}}$ iff $\mathcal{N} \vDash \mathbf{A}^{\mathsf{t}}_{\mathrm{R}}$, for any $\mathscr{L}$-sentence $\mathbf{A}^{\mathsf{t}}$. This completes the proof of Lemma 5 (and hence of the Conservativeness Theorem). □

D Richer conservative fictions

As mentioned in §4.3, we could have used a richer theory than NAT, and still obtained a Conservativeness Theorem. For example, consider these three schemes:

(a) $\exists(app_n(u^{\mathsf{e}}, x_1^{\mathsf{e}}, \ldots, x_n^{\mathsf{e}})) \rightarrow \neg real(u)$
(b) $\exists(app_n(\underline{u^{\alpha_1(\alpha_2(\ldots(\alpha_n\mathsf{t})\ldots))}}, x_1^{\mathsf{e}}, \ldots, x_n^{\mathsf{e}})) \rightarrow (\exists v_1^{\alpha_1} x_1 = \underline{v_1} \wedge \ldots \wedge \exists v_n^{\alpha_n} x_n = \underline{v_n})$
(c) $(real(u^{\alpha\beta}) \wedge real(v^{\alpha\beta}) \wedge \forall x^\alpha(real(x) \rightarrow u(x) \simeq_\beta v(x))) \rightarrow u =_{\alpha\beta} v$

The first says that no real object applies to any objects; this is sensible, since real objects are not nominalizations of higher-order entities. (The same principle prevents anything from instantiating Socrates.) The second ensures that meaningful applications really must keep track of the behaviour of higher-order entities. (The same principle prevents Socrates from instantiating $\underline{virtue^{(\mathsf{et})\mathsf{t}}}$.) The third says that real entities which agree on all real inputs are identical (cf. *Step 5* Fact 2, above). All three schemes can be added to NAT, and the result would still be conservative: this is immediate from our proof of the Conservativeness Theorem, since $\mathcal{N}$, as constructed, satisfies these schemes.

But there is no need to stop just with principles which we can read off from $\mathcal{N}$; we can also tweak the construction a little. For example: in constructing the structure, we might decide to add fictitious entities, which enable us to 'tag' each object which is the nominalization of some type α-entity with a label indicating that this is so. To do this, we would need a formal theory of such tags—some theory of syntax would do—and then we would add some scheme like:

(d) $tag(x^{\mathsf{e}}) =_{\mathsf{e}} \ulcorner\alpha\urcorner \leftrightarrow \exists v^\alpha(real(v) \wedge x = \underline{v})$ all $\alpha \neq \mathsf{e}$.

We could then go on to formulate principles concerning the higher-order origins of nominalizations, within the fictional object language.[75] To ensure the result is conservative, we just need to take suitable care in constructing $\mathcal{N}$; but this is straightforward, given what has come before, and we leave the details to enthusiastic readers who want to put the fiction to even greater work.

[75] Cf. Hale and Linnebo's (2020: 102–3) schemes (Bridge-*OC*) and (5.1).

E Bridge-principles

As discussed in §7.2, we can also add bridge-principles conservatively. Let $\mathtt{r}_n$ be the type of an n-place first-level relation, i.e. $\mathtt{e}(\ldots(\mathtt{et})\ldots)$, with n-occurrences of 'e'. Then we can restate the definition of a bridge-principle, which we gave in §7.2, to include a useful index, **P**:

Definition 6: A *bridge-principle*, **P**, for a theory T is a $\mathrm{wff}_{\mathtt{t}}$ of this form:

$$\left(real(u_1^{\alpha_1^{\mathbf{P}}}) \wedge \ldots \wedge real(u_n^{\alpha_n^{\mathbf{P}}})\right) \rightarrow \mathbf{A}_{\mathbf{P}}^{\mathtt{r}_n}(\underline{u_1}, \ldots, \underline{u_n}) =_{\mathtt{t}} \mathbf{B}_{\mathbf{P}}^{\alpha_1^{\mathbf{P}}(\ldots(\alpha_n^{\mathbf{P}}\mathtt{t})\ldots)}(u_1, \ldots, u_n)$$

where $\mathbf{A_P}$ and $\mathbf{B_P}$ are T-constants and $\alpha_i^{\mathbf{P}} \neq \mathtt{e}$ for some $1 \leq i \leq n$.

The intuitive idea behind conservative addition of bridge-principles is this: bridge-principles which cannot conflict with each other cannot disrupt T_{U}. Here is a precise way to spell out the impossibility of conflict:

Definition 7: Where Δ is a set of bridge-principles for T, say that Δ is *T-friendly* iff T proves every grammatical instance of this scheme, for all $\mathbf{P}, \mathbf{Q} \in \Delta$:

$$\mathbf{A_P}(u_1^{\mathtt{e}}, \ldots, u_i^{\mathtt{e}}) = \mathbf{A_Q}(v_1^{\mathtt{e}}, \ldots, v_j^{\mathtt{e}}) \rightarrow \mathbf{B_P}(u_1, \ldots, u_i) = \mathbf{B_Q}(v_1, \ldots, v_j)$$

Note that, for example, we allow instances where $\mathbf{A_P}$ has type $\mathtt{r}_n$ and $i < n$; then $\mathbf{A_P}(u_1^{\mathtt{e}}, \ldots, u_i^{\mathtt{e}})$ has type $\mathtt{r}_{n-i}$.

The reader can confirm that Definition 7 covers the intuitive cases of possible conflict which we raised in §7.2. We can now prove the following strengthening of our original Conservativeness Theorem:

Bridged Conservativeness Theorem: Let T be a PFTT theory in some signature, $\mathscr{L}$, which is disjoint from NAT's non-logical vocabulary. Let Δ be a set of T-friendly bridge-principles. If $T_{\mathrm{U}} \cup \Delta \vdash \mathbf{A}_{\mathrm{R}}$ then $T \vdash \mathbf{A}$, for any $\mathscr{L}$-sentence $\mathbf{A}^{\mathtt{t}}$.

The proof strategy is exactly as for the original Conservativeness Theorem in §C: we show how to 'expand' any model $\mathcal{M}$ of T into a model $\mathcal{N}$ of $T_{\mathrm{U}} \cup \mathrm{NAT} \cup \Delta$. The only part of our original proof which needs adjustment is *Step 2*, where we define $\star$ and the R_αs; these definitions must be tweaked, to ensure that $\mathcal{N} \vDash \Delta$. The remainder of this appendix spells out that tweak.

For each type which is *not* some $\mathtt{r}_n$, we define $\star$ and R_α exactly as in *Step 2* of §C. We provide special treatment, though, for the $\mathtt{r}_n$s. For each $i \leq n \in \mathbb{N}$, each $a \in M_{\mathtt{r}_n}$, all $x_1, \ldots, x_{n-i} \in M_{\mathtt{e}}$, and all $y_1, \ldots, y_i$, say:[76]

[76] We use $\vec{}$ to abbreviate lists, allowing context to indicate their length. So $a_\star(\vec{x}, \vec{y})$ abbreviates $a_\star(x_1, \ldots, x_{n-i}, y_1, \ldots, y_i)$.

$$a_\star(\vec{x},\vec{y}) := \begin{cases} a(\vec{x},\vec{y}) & \text{if } \vec{y} \in M_{\mathsf{e}} \quad (1) \\ \mathbf{B}_{\mathbf{P}}^{\mathcal{M}}(\vec{z},\vec{s}) & \text{if } \mathbf{P} \in \Delta, \text{ and } y_k = \mu_{\alpha_{m-i+k}}(s_k) \text{ for all } 1 \leq k \leq i, \text{ and} \\ & \quad z_1,\ldots,z_{m-i} \in M_{\mathsf{e}}, \text{ and } a(\vec{x}) = (\mathbf{A}_{\mathbf{P}}^{\mathsf{r}_m})^{\mathcal{M}}(\vec{z}) \quad (2) \\ \mathrm{F} & \text{otherwise} \quad (3) \end{cases}$$

Then, as before, we define $R_{\mathsf{r}_n} := \{a_\star : a \in M_{\mathsf{r}_n}\}$.

We must confirm that $\star$ is well-defined. First, observe that clauses (1) and (2) cannot conflict. Assume clause (2) applies. Then $\alpha^{\mathbf{P}}_{m-i+k} \neq \mathsf{e}$ for some $1 \leq k \leq i$ by Definition 6, so that $y_k = \mu_{\alpha^{\mathbf{P}}_{m-i+k}}(s_k) \notin M_{\mathsf{e}}$, as the U_αs are disjoint (see *Step 2* of §C). So clause (1) does not apply.

To complete the demonstration that $\star$ is well-defined, it suffices to show that clause (2) never causes conflict. So suppose that $a_\star(\vec{x},\vec{y})$ is to be defined using clause (2) and two witnessing bridge-principles $\mathbf{P},\mathbf{Q} \in \Delta$. So we have $\vec{u},\vec{v} \in M_{\mathsf{e}}$, and $a(\vec{x}) = (\mathbf{A}_{\mathbf{P}}^{\mathsf{r}_l})^{\mathcal{M}}(\vec{u}) = (\mathbf{A}_{\mathbf{Q}}^{\mathsf{r}_m})^{\mathcal{M}}(\vec{v})$, and we are to assign both:

$$a_\star(\vec{x},\vec{y}) := \mathbf{B}_{\mathbf{P}}^{\mathcal{M}}(\vec{u},\vec{s}) \qquad \text{where } y_k = \mu_{\alpha^{\mathbf{P}}_{l-i+k}}(s_k) \text{ for all } 1 \leq k \leq i$$

$$a_\star(\vec{x},\vec{y}) := \mathbf{B}_{\mathbf{Q}}^{\mathcal{M}}(\vec{v},\vec{t}) \qquad \text{where } y_k = \mu_{\alpha^{\mathbf{Q}}_{m-i+k}}(t_k) \text{ for all } 1 \leq k \leq i$$

Since $\mathcal{M} \vDash T$ and Δ is T-friendly, $\mathbf{B}_{\mathbf{P}}^{\mathcal{M}}(\vec{u}) = \mathbf{B}_{\mathbf{Q}}^{\mathcal{M}}(\vec{v})$. For each $1 \leq k \leq i$, we have: $y_k = \mu_{\alpha^{\mathbf{P}}_{l-i+k}}(s_k) = \mu_{\alpha^{\mathbf{Q}}_{m-i+k}}(t_k)$, so that $s_k = t_k$ by the bijectivity of the U_αs. Hence $\mathbf{B}_{\mathbf{P}}^{\mathcal{M}}(\vec{u},\vec{s}) = \mathbf{B}_{\mathbf{Q}}^{\mathcal{M}}(\vec{v},\vec{t})$. This completes the proof that $\star$ is well-defined.

We now check that $\star$, as redefined, still possesses the key properties which we invoked in *Steps 3–5* of §C.

First: restricting $\star$ provides a *bijection* $M_\alpha \longrightarrow R_\alpha$. When $\alpha = \mathsf{r}_n$ for some n, this is immediate from clause (1); otherwise, this holds as in §C.

Second: $\star$ is *distributive*, in that $a_\star(x_\star) = (a(x))_\star$, whenever $a(x)$ is defined. If $a \notin M_{\mathsf{r}_n}$ for any n, this holds by definition (see *Step 2* of §C); so it suffices to consider $a \in M_{\mathsf{r}_n}$ and $x \in M_{\mathsf{e}}$. Since then $x_\star = x$, it suffices to show that $a_\star(x) = (a(x))_\star$. For readability, let $b = a(x)$; we want to show that $a_\star(x,\vec{y}) = (a_\star(x))(\vec{y}) = b_\star(\vec{y})$ for all $\vec{y} \in N_{\mathsf{e}}$. There are three cases to consider, corresponding to the three clauses we used to define $\star$:

Case (1). We have some $\vec{y} \in M_{\mathsf{e}}$ such that $a_\star(x,\vec{y}) = a(x,\vec{y}) = b(\vec{y})$; now also $b_\star(\vec{y}) = b(\vec{y})$ by clause (1).

Case (2). We have some $\mathbf{P} \in \Delta$ and $\vec{z} \in M_{\mathsf{e}}$ such that, relabelling $\vec{y}$ as $\vec{u},\vec{v}$ and where each $v_k = \mu_{\alpha_{m-i+k}}(s_k)$, we have $a_\star(x,\vec{y}) = a_\star(x,\vec{u},\vec{v}) = \mathbf{B}_{\mathbf{P}}^{\mathcal{M}}(\vec{z},\vec{s})$ and $a(x,\vec{u}) = (\mathbf{A}_{\mathbf{P}}^{\mathsf{r}_m})^{\mathcal{M}}(\vec{z})$. Since $b(\vec{u}) = (\mathbf{A}_{\mathbf{P}}^{\mathsf{r}_m})^{\mathcal{M}}(\vec{z})$, now also $b_\star(\vec{y}) = b_\star(\vec{u},\vec{v}) = \mathbf{B}_{\mathbf{P}}^{\mathcal{M}}(\vec{z},\vec{s})$ by clause (2).

Case (3). Similarly, $b_\star(\vec{y}) = b(\vec{y}) = \mathrm{F}$ by clause (3).

This shows that $\star$, as redefined, still possesses the key properties which we invoked in *Steps 3–5* of §C. It follows, as before, that $\mathcal{N}$ satisfies $T_{\mathrm{U}} \cup \mathrm{NAT}$.

It remains to show that $\mathcal{N} \vDash \Delta$. Fix any $\mathbf{P} \in \Delta$. Recalling that $\mathbf{A_P}$ and $\mathbf{B_P}$ are both $\mathtt{t}$-kind, as in *Step 5* both $(\mathbf{A}_{\mathbf{P}}^{\mathcal{M}})_\star = \mathbf{A}_{\mathbf{P}}^{\mathcal{N}}$ and $(\mathbf{B}_{\mathbf{P}}^{\mathcal{M}})_\star = \mathbf{B}_{\mathbf{P}}^{\mathcal{N}}$. Fix $x_1 \in R_{\alpha_1},\ldots,x_n \in R_{\alpha_n}$;

by the properties of $\star$, for each $1 \leq i \leq n$ there is some unique $s_i \in M_{\alpha_i}$ such that $x_i = (s_i)_\star$, and moreover $\underline{x_i} = \mu_{\alpha_i}(s_i)$. Now $\mathcal{N} \vDash \mathbf{P}$ as:

$$\begin{aligned}
\mathbf{A}_{\mathbf{P}}^{\mathcal{N}}(\underline{x_1}, \ldots, \underline{x_n}) &= (\mathbf{A}_{\mathbf{P}}^{\mathcal{M}})_\star(\mu_{\alpha_1}(s_1), \ldots, \mu_{\alpha_n}(s_n)) \\
&= \mathbf{B}_{\mathbf{P}}^{\mathcal{M}}(s_1, \ldots, s_n) \\
&= (\mathbf{B}_{\mathbf{P}}^{\mathcal{N}}(s_1, \ldots, s_n))_\star \\
&= (\mathbf{B}_{\mathbf{P}}^{\mathcal{M}})_\star((s_1)_\star, \ldots, (s_n)_\star) \\
&= \mathbf{B}_{\mathbf{P}}^{\mathcal{N}}(x_1, \ldots, x_n)
\end{aligned}$$

The second equality invokes clause (2), letting $m = i = n$; the third equality is as in *Step 5*; and the fourth equality uses $\star$'s distributivity.

References

Andrews, Peter B (2002). *An Introduction to Mathematical Logic and Type Theory: To truth through proof*. 2nd ed. Springer.

Armstrong, David (1978). *Nominalism and Realism*. Vol. 1. Universals & Scientific Realism. Cambridge: Cambridge University Press.

— (1980). 'Against ostrich nominalism: A reply to Michael Devitt'. *Pacific Philosophical Quarterly* 61, pp.440–9.

— (1989). *Universals: An Opinionated Introduction*. Boulder, CO: Westview Press.

— (2004). *A World of States of Affairs*. Cambridge: Cambridge University Press.

Balaguer, Mark (1996). 'A fictionalist account of the indispensable applications of mathematics'. *Philosophical Studies* 83, pp.291–314.

— (1998). *Platonism and Anti-Platonism in Mathematics*. Oxford: Oxford University Press.

Båve, Arvid (2015). 'A deflationist error theory of properties'. *Dialectica* 69, pp.23–59.

Beaney, Michael, ed. (1997). *The Frege Reader*. Oxford: Blackwell.

Blackburn, Simon (1987). 'Morals and modals'. In: *Essays in Quasi-Realism* (1993). Oxford: Oxford University Press, pp.52–74.

— (2005). 'Quasi-realism no fictionalism'. In: *Fictionalism in Metaphysics*. Ed. by Mark Eli Kalderon. Oxford: Oxford University Press, pp.322–38.

Burgess, John P. (2004). 'Mathematics and *Bleak House*'. *Philosophia Mathematica* 12, pp.18–36.

Burgess, John P. and Gideon Rosen (1997). *A Subject with No Object: Strategies for Nominalistic Interpretation of Mathematics*. Oxford: Oxford University Press.

Button, Tim and Robert Trueman (2022). 'Against cumulative type theory'. *Review of Symbolic Logic* 15.4, pp.907–49.

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

Chierchia, Gennaro (1982). 'Nominalization and Montague grammar: A semantics without types for natural languages'. *Linguistics and Philosophy* 5.3, pp.303–54.

— (1984). 'Topics in the Syntax and Semantics of Infinitives and Gerunds'. PhD thesis. UMass Amherst.

— (1985). 'Formal semantics and the grammar of predication'. *Linguistic Inquiry* 16.3, pp.417–443.

Chierchia, Gennaro and Raymond Turner (1988). 'Semantics and property theory'. *Linguistics and Philosophy* 11.3, pp.261–302.
Church, Alonzo (1940). 'A formulation of the simple theory of types'. *The Journal of Symbolic Logic* 5.2, pp.56–68.
Cocchiarella, Nino B (1972). 'Properties as individuals in formal ontology'. *Noûs* 6.2, pp.165–87.
— (1974). 'Fregean semantics for a realist ontology'. *Notre Dame Journal of Formal Logic* 15.4, pp.552–68.
— (1975a). 'Second-order theories of predication: Old and new foundations'. *Noûs* 9.1, pp.33–53.
— (1975b). 'The logic of nominalized predicates and its philosophical interpretation'. *Erkenntnis* 13.3, pp.339–69.
Cresswell, M J (1973). *Logics and Languages*. London: Methuen.
Dorr, Cian (2016). 'To be F is to be G'. *Philosophical Perspectives* 30, pp.39–134.
Farmer, William M (1990). 'A partial functions version of Church's simply theory of types'. *The Journal of Symbolic Logic* 55.3, pp.1269–91.
Field, Hartry (1989/1991). *Relaism, Mathematics and Modality*. Oxford: Blackwell.
— (1980/2016). *Science Without Numbers*. 2nd ed. Oxford: Oxford University Press.
Florio, Salvatore and Nicholas K Jones (2019). 'Unrestricted quantification and the structure of type theory'. *Philosophy and Phenomenological Research*, pp.1–21.
Frege, Gottlob (1892). 'On concept and object'. In: Beaney 1997, pp.181–93.
— (1891–5). 'Comments on "Sense and reference"'. In: Beaney 1997, pp.172–80.
— (1924/5). 'Sources of knowledge of mathematics and the mathematical natural sciences'. In: *Posthumous Writings*. Ed. by Hans Hermes et al. Trans. by Peter Long and Roger White. Oxford: Blackwell, pp.267–74.
Gaskin, Richard (1995). 'Bradley's Regress, the copula and the unity of the proposition'. *The Philosophical Quarterly* 45, pp.161–80.
— (2008). *The Unity of the Proposition*. Oxford: Oxford University Press.
Geach, Peter (1976). 'Saying and showing in Frege and Wittgenstein'. *Acta Philosophica Fennica* 28, pp.54–70.
Goodman, Jeremy (2017). 'Reality is not structured'. *Analysis* 77, pp.43–53.
Hale, Bob (2010). 'The bearable lightness of being'. *Axiomathes* 20, pp.399–422.
— (2013). *Necessary Beings*. Oxford: Oxford University Press.
Hale, Bob and Øystein Linnebo (2020). 'Ontological categories and the problem of expressibility'. In: *Essence and Existence: Selected Essays*. Ed. by Jessica Leech. Oxford: Oxford University Press, pp.73–103.
Hale, Bob and Crispin Wright (2012). 'Horse sense'. *The Journal of Philosophy* 109, pp.85–131.
Heim, Irene and Angelika Kratzer (1998). *Semantics in Generative Grammar*. Oxford: Blackwell.
Hofweber, Thomas (2018). 'Replies to Eklund and Uzquiano'. *Analysis* 78, pp.315–34.
Jackson, Frank (1977). 'Statements about universals'. *Mind* 86, pp.427–9.
Jones, Nicholas K (2016). 'A higher-order solution to the problem of the concept *horse*'. *Ergo* 3, pp.132–66.
— (2018). 'Nominalist realism'. *Noûs* 52, pp.808–35.
King, Jeffrey et al. (2014). *New Thinking about Propositions*. Oxford: Oxford University Press.

Krämer, Stephan (2014). 'Semantic values in higher-order semantics'. *Philosophical Studies* 168, pp.709–24.
Künne, Wolfgang (2003). *Conceptions of Truth*. Oxford: Oxford University Press.
Leng, Mary (2010). *Mathematics and Reality*. Oxford: Oxford University Press.
Liebesman, David (2015). 'Predication as ascription'. *Mind* 124, pp.517–69.
Linnebo, Øystein (2006). 'Sets, properties, and unrestricted quantification'. In: *Unrestricted Quantification*. Ed. by Agustín Rayo and Gabriel Uzquiano. Oxford: Oxford University Press, pp.149–178.
— (forthcoming). 'Higher-order comprehension and the possibility of fine structure'. In: *this volume*.
Linnebo, Øystein and Agustín Rayo (2012). 'Hierarchies ontological and ideological'. *Mind* 121.482, pp.269–308.
MacBride, Fraser (2011). 'Impure reference: A way around the concept *horse* paradox'. *Philosophical Perpectives* 25, pp.297–312.
Magidor, Ofra (2009). 'The last dogma of type confusions'. *Proceedings of the Aristotelian Society* 109, pp.1–29.
Melia, Joseph (2000). 'Weasling away the indispensability argument'. *Mind* 109, pp.455–80.
Menzel, Christopher (forthcoming). 'Pure logic: First-order, type-free, hyperintensional'. In: *this volume*.
Montague, Richard (1973). 'The proper treatment of quantification in ordinary English'. In: *Portner and Partee 2002*, pp.17–34.
Noonan, Harold (2006). 'The concept horse'. In: *Universals, Concepts and Qualities*. Ed. by Peter Strawson and Arindam Chakrabarti. Aldershot: Ashgate, pp.155–76.
Pap, Arthur (1959). 'Nominalism, empiricism and universals I'. *The Philosophical Quarterly* 9, pp.330–40.
Parsons, Terence (1979). 'The theory of types and ordinary language'. In: *Linguistics, Philosophy and Montague Grammar*. Ed. by Steven Davis and Marianne Mithun. Austin: University of Texas Press, pp.127–51.
Partee, Barbara (1986). 'Noun-phrase interpretation and type-shifting principles'. In: *Portner and Partee 2002*, pp.357–81.
Portner, Paul and Barbara Partee, eds. (2002). *Formal Semantics: The essential readings*. Oxford: Blackwell.
Price, Michael (2016). 'Naming the concept *horse*'. *Philosophical Studies* 173, pp.2727–43.
Prior, Arthur (1971). *Objects of Thought*. Oxford: Oxford University Press.
Proops, Ian (2013). 'What is Frege's 'concept *horse* problem'?' In: *Wittgenstein's Tractatus: History and Interpretation*. Ed. by Peter Sullivan and Michael Potter. Oxford: Oxford University Press, pp.76–96.
Rayo, Agustín and Stephen Yablo (2001). 'Nominalism through de-nominalization'. *Noûs* 35, pp.74–92.
Rieppel, Michael (2016). 'Being something: Properties and predicative quantification'. *Mind* 125.499, pp.643–89.
— (2018). 'Denoting and disquoting'. *Australasian Journal of Philosophy* 96, pp.548–61.
Rodriguez-Pereyra, Gonzalo (2000). 'What is the problem of universals?' *Mind* 109, pp.255–74.
Rosefeldt, Tobias (2008). '"That"-clauses and non-nominal quantification'. *Philosophical Studies* 137.

Russell, Bertrand (1903). *The Principles of Mathematics*. London: Allen and Unwin.
Schiffer, Stephen (2003). *The Things We Mean*. Oxford: Oxford University Press.
Schiller, Francis Canning Scott (1912). 'Critical notice of Vaihinger (1911)'. *Mind* 21.81, pp.93–104.
Skiba, Lukas (2020). 'Higher-order metaphysics and the tropes vs. universals dispute'. *Philosophical Studies*.
Stanley, Jason (2001). 'Hermeneutic fictionalism'. *Midwest Studies in Philosophy* 25, pp.36–71.
Thomasson, Amie (2013). 'Fictionalism versus deflationism'. *Mind* 122.488, pp.1023–51.
— (2014). *Ontology Made Easy*. Oxford: Oxford University Press.
Tichý, Pavel (1988). *The Foundations of Frege's Logic*. Berlin: Walter de Gruyter.
Trueman, Robert (2015). 'The concept *horse* with no name'. *Philosophical Studies* 172, pp.1889–906.
— (2018). 'The Prenective View of propositional content'. *Synthese* 195, pp.1799–1825.
— (2020). 'Idealism and the identity theory of truth'. *Mind*, pp.1–25.
— (2021). *Properties and Propositions: The Metaphysics of Higher-Order Logic*. Cambridge: Cambridge University Press.
Vaihinger, Hans (1911). *Die Philosophie des Als Ob. System der Theoretischen, Praktischen und Religiosen Fiktionen der Menschheit auf Grund Eines Idealistischen Positivismus, mit einem Anhang uber Kant und Nietzsche*. Berlin: Reuther & Reichard.
Williamson, Timothy (2003). 'Everything'. *Philosophical Perspectives* 17.1, pp.415–65.
— (2013). *Modal Logic as Metaphysics*. Oxford: Oxford University Press.
Wright, Crispin (1998). 'Why Frege does not deserve his grain of salt'. In: *The Reason's Proper Study*. Ed. by Bob Hale and Crispin Wright. Oxford: Oxford University Press (2001), pp.72–90.
Yablo, Stephen (2005). 'The myth of the seven'. In: *Fictionalism in Metaphysics*. Ed. by Mark Kalderon. Oxford: Oxford University Press, pp.88–115.