Definition by Proxy

Samuel Z. Elgin

"It is a philosophical problem how there can ever be unobvious analyticity...it is everybody's problem and it is not to be solved by denying the phenomenon."—David Lewis[1]

Abstract

I take some initial steps toward a theory of real definition, drawing upon recent developments in higher-order logic. The resulting account allows for extremely fine-grained distinctions (it can distinguish between any relata that differ in their syntactic structure, while avoiding the Russell-Myhill problem). It is the first account that can consistently embrace three desirable logical principles that initially appear to be incompatible: the *Identification Hypothesis* (if F is, by definition, G then there is a sense in which F is the same as G), *Irreflexivity* (there are no reflexive definitions) and *Leibniz's Law*. Additionally, it possesses the resources needed to resolve the paradox of analysis.

Three Puzzles of Definition

The notion of definition has occupied a central role in philosophy since the inception of our discipline—and in no field is its significance more apparent than metaphysics. Putative examples include the following:[2]

1. To be just is, by definition, for each part of one's soul to do its proper work.
2. To know that p is, by definition, to have a justified true belief that p.
3. To be even is, by definition, to be a natural number divisible by two without remainder.
4. To be made of water is, by definition, to be made of the chemical compound H_2O.
5. To be a moral agent is, by definition, to be bound by the categorical imperative.

In these cases, the entities being defined are properties and relations themselves—rather than the words that denote them. Our concern is with the definition of *being a moral agent*,

[1]I am deeply indebted to Cian Dorr, who first suggested to me that we might understand definition in terms of Fritz-style proxies. I also thank Michael Della Rocca, Catherine Elgin, Peter Fritz, Joseph Martinez, On Yi sin, Joseph Stratmann, eric Watkins and Isaac Wilhelmfor their helpful comments on earlier versions of this paper.

[2]It is far from clear that those who advanced accounts along these lines intended them to be definitions. I suspect that many philosophers have not taken a firm stance on the status of the accounts that they defend. Nevertheless, I also suspect that many can be reasonably read as putative definitions—regardless of whether that was the original intent.

rather than 'being a moral agent.' And so, the definitions at issue are not nominal, but real.[3]

Definition gives rise to a number of puzzles that are not altogether easy to resolve. There is, firstly, a paradox of analysis. It is unclear how any account could be both substantive and true.[4] For an analysis to be substantive, its content its content must convey information that its object does not.[5] But if an object and content differ informationally, then the object is not the same as its content, and so the putative analysis is false. Utilitarianism is undermined neither because it ignores the distinction between persons, nor by the plethora of our acts' unforeseen consequences, but rather by the mere fact that *is morally right* differs in significance from *maximizes utility.* And that old Aristotelian theory that to be human is, by definition, to be a rational animal is falsified neither by the theory of evolution nor by the systematic irrationality of our species, but rather by the fact that *is human* differs in significance from *is a rational animal.* The paradox suggests that the quest for analyses is doomed from the outset, for an account's very substance undermines its truth.

Another—and arguably intimately related—puzzle concerns the logic of definition.[6] There are three principles governing this logic that seem indispensable, yet also incompatible:

Identification Hypothesis	If F is, by definition, G then there is an important sense in which F is the same as G.
Leibniz's Law	Terms that denote the same thing may be substituted for one another.
Irreflexivity	There are no reflexive definitions.

It is worth clarifying what these principles mean at the outset. The *Identification Hypothesis* concerns numerical, higher-order identity. Various locutions express identities

[3]Throughout this paper, 'definition' refers to real definition unless otherwise specified. I myself maintain that nominal definitions are a special sort of real definition; they are the real definitions of linguistic items like words, clauses and phrases. In particular, I hold that these definitions are given by the semantic content of the linguistic item being defined. However, nothing in this paper depends on that conception of nominal definition.

[4]Here, my use of 'analysis,' 'account,' and 'definition' are all interchangeable—employed only for ease of prose. Though there may well be examples of philosophical analyses or accounts that are not definitions, these are not examples that I discuss in this paper.

[5]Here, 'object' and 'content' refer to the property, proposition or relation that is being defined—and to that which the property, proposition or relation is defined in terms of—respectively. For example, in the sentence 'To be a bachelor is, by definition, to be an unmarried male,' *is a bachelor* is the object of analysis while *is an unmarried male* is the content. The object of definition as sometimes also referred to as 'the definiendum' and the content as 'the definiens.'

[6]I do not wish to engage in a lengthy discussion about the relation between the various puzzles. Perhaps some maintain that the first two are different ways of framing the same problem—in which case this theory of definition resolves two problems, not three. But I suspect that some see daylight between these problems. As is made clear below, the second is a purely logical problem—while the first concerns the ability to extract information from the content of analysis that cannot be extracted from the object.

of this sort; 'To be F is to be G,' 'Being F is the same as being G,' and 'To be F just is to be G' are all natural contenders.[7] Importantly, the identities at issue are not merely qualitative. If F is, by definition, G, then it is not only the case that F and G are similar in certain regards; rather, there is a sense in which they are one and the same. Additionally, *Leibniz's Law* is given as a substitution principle—while some may be more accustomed to a principle concerning properties (along the lines of 'If a is identical to b, then a and b bear the same properties'). Note that it is possible to generate the property-theoretic version from the substitution principle; if a is identical to b, then the substitution principle allows the inference from Fa to Fb. Given the higher-order framework to come, it will be more useful to employ the substitution principle—one with applications for terms other than first-order properties.

Each of these principles has been assumed to be true without argument. That is to say, metaphysicians not only have taken them to have some measure of intuitive appeal, but rather have held that they are *so* intuitive that no defense need be given. Indeed, each principle has been treated as a compulsory starting-point in a theory of definition; some have maintained that an account that falsifies one is to be rejected solely for that reason.[8] But this is not to say that no argument can be given on their behalf. On the contrary, there are considerations that lend support to each.

Many maintain that definition is reductive. If {Socrates} is, by definition, the set containing only Socrates then {Socrates} is nothing more than the set containing only Socrates—and if the number two is, by definition, the successor to the number one, then the number two is nothing more than the successor to the number one. Any theory of definition that denies the *Identification Hypothesis* appears reductive in name only (if that). So to account for the reductive aspect of definition, we ought to maintain that the *Identification Hypothesis* is true.

Leibniz's Law is also extraordinarily intuitive.[9] If Cicero is an orator and Cicero is identical to Tully, then Tully is an orator—and if Hesperus appears in the evening sky and Hesperus is identical to Phosphorus, then Phosphorus appears in the evening sky. Admittedly, there are canonical challenges for *Leibniz's Law* arising from opaque predicates like 'believes,' but many continue to endorse the view that co-referring terms can be substituted for one another *salva veritate*—difficult cases notwithstanding. It may even be that *Leibniz's Law* is partially constitutive of what we mean by 'identity': it is difficult to even understand what it would be for identical objects to not be substitutable for one another.

[7] The literature on higher-order identification has exploded in recent years. See, for example, Rayo (2013); Dorr (2016); Correia and Skiles (2019); Bacon and Russell (2019); Bacon (2019); Caie, Goodman and Lederman (2020).

[8] See, e.g., Correia (2017) for someone who treats the *Identification Hypothesis* and *Irreflexivity* as compulsory starting-points, and Dorr (2016) for someone who endorses *Leibniz's Law* without argument.

[9] There has been sustained discussion of higher-order systems that reject Leibniz's Law—see Caie, Goodman and Lederman (2020); Bacon and Russell (2019); Bacon (2019). Those who would reject Leibniz's Law in this context might appeal to the systems that they develop. I do not do so, as my account is one on which Leibniz's Law is true.

Irreflexivity, in turn, reflects the thought that definition tracks relative fundamentality (on at least one conception of fundamentality). If water is, by definition, the chemical compound H_2O, then hydrogen and oxygen are more fundamental than water—and if hydrogen is, by definition, the element with a single proton, then protons are more fundamental than hydrogen. A bit more precisely, we might maintain that if B occurs within the content of the definition of A, then B is more fundamental than A.[10] Given the (not unreasonable) assumption that nothing is more fundamental than itself, it follows that there are no reflexive definitions.

But despite their initial appeal and stalwart motivation, it seems that at least one of these principles must be rejected, for the three appear to be incompatible.[11] Take an arbitrary F and G such that F is, by definition, G. The conflict is generated in the following way:

i)	$Def(F,G)$	Supposition
ii)	$F = G$	*i*, Identification Hypothesis
iii)	$Def(F,F)$	*i*, *ii*, Leibniz's Law
iv)	$\bot$	*iii*, Irreflexivity

Less formally, if F is, by definition G, then—by the *Identification Hypothesis*—F is the same as G.[12] Because F is the same as G, *Leibniz's Law* entails that 'F' and 'G' may be substituted for one another in every context—one of which results in 'F is, by definition, F'—contra *Irreflexivity*. For this reason, philosophers in this area fall into one of three camps: those who reject the *Identification Hypothesis* (e.g., Fine (1994); Rosen (2015)), those who reject *Leibniz's Law* (e.g., Correia (2017)) and those who reject *Irreflexivity* (e.g., Dorr (2016)).[13] There is presently no theory of definition that validates all three principles.

[10]See Fine (1994, 1995) for someone who defends this notion of relative fundamentality (in his words, 'ontological dependence').

[11]A bit more precisely, they are apparently incompatible on the assumption that there is at least one definition. If 'definition' had an empty extension, then the *Identification Hypothesis* and *Irreflexivity* would both be vacuously true—and there is no reason to suspect that definition would pose problems for *Leibniz's Law* that it did not already face. This assumption, though strictly needed to derive the conflict, is often overlooked in the literature.

[12]In this description—and in the previous quasi-formal derivation—the Identification Hypothesis is interpreted as the claim that if F is, by definition, G, then F is literally identical to G. While this is a natural interpretation, it need not be the only one. The view I develop preserves a relation extremely close to identity (while not strictly being identity)—and this is ultimately how it avoids the conflict at hand.

[13]Fine (1994)'s view is that the definition of an object is the collection of properties essential to that object. Many deny that objects are identical to collections of essential properties (though there may be versions of bundle-theory on which they are) so there is reason to think that his rejection of *Identification* is widespread. However, even on bundle-theoretic views, Fine allows for definitions that do not uniquely identify the object that they define. That is, there may be objects a and b such that $Def(a,F)$ and $Def(b,F)$ and $a \neq b$. For this reason, Fine's account allows for violations of *Identification*. Relatedly, Rosen does not explicitly deny the *Identification Hypothesis*, but—as Correia (2017) argues—it is possible to generate violations of *Identification* on his account. Additionally, Dorr (2016) does not strictly deny

Given the conflict, this seems to be no accident of the literature. No account *could* validate all three—on pain of inconsistency. And so every theory will have an undesirable logical attribute (though they vary with respect to which undesirable logical attribute they have).

A third puzzle concerns the granularity of propositions and relations being defined. Arguably, definition makes extremely fine-grained distinctions. It may be that 'To be a triangle is, by definition, to be a polygon with three angles' is true while 'To be a triangle is, by definition, to be a polygon with three sides' is false. If this is so, then an adequate account of definition must distinguish 'to be a polygon with three angles' from 'to be a polygon with three sides.' Treating properties as functions from possible worlds to extensions will not do—as the two presumably have the same extension in every possible world.

An obvious thought is to appeal to structured propositions—as they possess the requisite granularity. Perhaps propositions, properties and relations have a structure that resembles the syntactic structure of sentences and clauses; perhaps they are composed out of worldly items in much the way that sentences are composed of words.[14] In particular, it may be that *being a polygon with three angles* is composed of material concerning angles, while *being a polygon with three sides* is composed of material concerning sides. In virtue of their different composition, the two properties are distinct—their necessarily identical extensions notwithstanding—and so can stand in different relations from one another.

Accounts of structured propositions have recently come under sustained assault. One of their central tenets is that if the proposition that Fa is identical to the proposition that Gb then $a = b$ and $F = G$.[15] For example, if the proposition that John is a brother is identical to the proposition that John is a male sibling, then John is identical to John and *being a brother* is identical to *being a male sibling.* This tenet is radically incompatible with an orthodox principle of higher-order logic: propositional identity is preserved through β-conversion.[16] Consider the binary relation R of *being the same height as*; two people stand in this relation to one another just in case they are equally tall. β-equivalence entails that $\lambda x.Rxx(a) = \lambda x.Rxa(a)$. The proposition that Mary is the same height as herself is identical to the proposition that Mary is the same height as Mary. On the structured proposition view, this entails that Mary is identical to Mary—no problems there. It also

---

Irreflexivity, but rather claims that identity performs the work of philosophical analysis, and that identity is reflexive. As I interpret Dorr (and those like him) this is as close as can be reasonably expected of a denial of *Irreflexivity.*

[14] Rosen (2015), for example, assumes this to be true in developing his account of definition.

[15] The presentation of this problem follows the model given in Dorr (2016). Quite recently, Bacon (Forthcoming) has presented a theory of structured account immune to some of these concerns. In particular, he shifts from an account that mirrors syntactic structure to one that mirrors pictorial structure; on his view, it is not the case that we can recover a unique property from each proposition.

[16] There are two principles in this area which can easily be confused. The first is an inferential principle: if ϕ is the β-conversion of ψ, then ψ entails that ϕ. For example, this principle may be used to infer Fa from $\lambda x.Fx(a)$. This is relatively uncontroversial (though some may deny it for languages with opaque predicates like 'believes'). A more controversial principle—which is the one I am presently concerned with—is that β equivalent terms denote the same thing. So, not only may one infer Fa from $\lambda x.Fx(a)$, but the two denote the very same proposition.

entails that $\lambda x.Rxx = \lambda x.Rxa$; the property of being the same height as oneself is identical to the property of being the same height as Mary. This is obviously absurd—the two are not even coextensive.

A second—and at least equally troubling—problem for structured propositions concerns their cardinality. This problem was originally discovered by Russell (1903) and noticed (apparently independently) by Myhill (1958). Another core commitment of structured propositions is that any differences in the syntactic structure of language correspond to differences in proposition. The fact that 'Jack is next to Jill' differs syntactically from 'It is false that Jack isn't next to Jill' ensures that the two sentences express different propositions. A bit roughly, the problem is that for every collection of propositions, it is possible to construct a sentence asserting that precisely the elements of that collection are true. For this reason, there is a mapping from the powerset of propositions to a unique sentence (i.e., there is a mapping from each element of the powerset of propositions to a sentence asserting that the propositions within that element are true).[17] If each sentence were itself to correspond to a unique proposition, then there would be a mapping from every element of the powerset of propositions to a unique proposition. But Cantor's Theorem entails that there is no such mapping. For every set s, there is no mapping from every element of the powerset of s to a unique element of s. And so, it cannot be that every difference in syntactic structure corresponds to a difference in proposition.

The upshot is this: in order for definition to make the distinctions that metaphysicians often take it to make, its relata must be fine-grained. It would be natural to appeal to structured propositions, which are suitably fine-grained. However, there are strong reasons to reject the structured proposition view. Those who would place a theory of definition on firm foundations ought to look elsewhere.

In light of the abundance and severity of these problems, some may be tempted to abandon the notion of definition entirely: to treat it as little more than a theoretical relic from a confused and imprecise time. I think that this is premature. Recent developments in higher-order logic suggest that the last of these puzzles, at least, can be resolved. Fritz (2021) demonstrates that there are structures that can make the fine-grained distinctions structured propositions had been intended to make, while avoiding the problems that structured propositions face. A natural thought is that definition relates terms to these higher-order proxies for structured propositions, rather than to structured propositions themselves. Interestingly, this shift resolves not only the third dilemma, but the first two as well. It allows us to consistently embrace the *Identification Hypothesis, Leibniz's Law* and *Irreflexivity*—and explains the respect in which definitions are both substantive and true. But in order to understand these resolutions, it is first necessary to describe what these higher-order proxies consist of.

But before I introduce the notion of a proxy precisely, a brief note on the account

[17] The use of set-theory here is purely expository; it is possible to generate the Russell-Myhill problem without reference to sets.

that emerges. This theory of definition strictly takes no stand on the granularity of the propositions, properties and relations that are being defined; it is compatible both coarse and fine-grained accounts. (Though it is incompatible with the sorts of ultra-fine-grained accounts mentioned above. As has already been noted, these accounts are themselves inconsistent—and so are incompatible with absolutely everything, including this theory). To take just one example, it takes no stand on the principles of idempotence (that proposition p is identical to $p \wedge p$ or $p \vee p$).[18] There are, however, some occasions when granularity affects how formalisms are introduced; I will flag points in footnotes when they arise. Nevertheless, it is most natural to accept this view on coarse-grained accounts. Proxies perform the theoretical work of making fine-grained distinctions; so propositions need not. On this conception, definition is a fine-grained relation in a coarse-grained world.

Structure by Proxy

One way to frame the previous problems for structured propositions is this: it is impossible, given the proposition that Fa, to recover property F and object a—in the sense that there may be a distinct G and b such that $Fa = Gb$. For example, the properties $\lambda x.Rxx$, $\lambda x.Rxa$ and $\lambda x.Rax$ can all be seen as figuring within the proposition Raa, so one cannot recover 'the' property contained within this proposition. Accounts of propositions that depend upon the possibility of singular recovery—like the structured proposition view—are false. Nevertheless, there is a higher-order term from which it *is* possible to recover a unique F and a: the relation between properties and objects that has only $< F, a >$ in its extension—i.e., the relation that property F stands in to object a and that no other property stands in to any other object. This is not a structured proposition. After all, it isn't a proposition of any kind. It is a relation between properties and objects, and is therefore not truth-evaluable. But precisely because it is a term that allows for the recovery of F and a, it may be used as a proxy for the structured proposition that Fa—and perform some of the theoretical work that metaphysicians naïvely believed structured propositions perform.[19]

Defining these proxies precisely requires a language to express them in. Let us adopt a simple, typed higher-order language L. Within L, there are two basic types—a type e for entities and a type t for sentences. Additionally, for any types τ_1, τ_2, $(\tau_1 \to \tau_2)$ is a type; nothing else is a type. In this language, the $\neg$ operator can be identified with a term of type $(t \to t)$. It is a function with sentences as inputs and sentences as outputs—in

[18] Many accounts of propositions hold that these are identical; for views that do not, see Dorr (2016); Elgin (Forthcoming).

[19] Conversationally, some have questioned the use of the phrase 'proxy for the structured proposition that Fa' on the grounds that proxies are only proxies for things that exist and—given the preceding arguments—structured propositions do not exist. Such philosophers are free to replace 'proxy for the structured proposition that Fa' with 'proxy for whatever fine-grained term allows for the recovery of F and a.' For ease of prose, I will continue to use phrases like 'proxies for structured propositions' throughout this paper—but similar substitutions could be made in all such cases if the reader desires.

particular, the output is the negation of the input. Similarly, the binary operators $\wedge, \vee$ can both be identified as terms of type $(t \to (t \to t))$. Monadic first-order predicates are identified with terms of type $(e \to t)$, dyadic first-order predicates are identified with terms of type $(e \to (e \to t))$, etc. There are also infinitely many variables of every type, as well as the corresponding λ abstracts needed to bind them. It would be natural, in this sort of language, to introduce quantifiers $\exists, \forall$ as terms of type $((\tau \to t) \to t)$ for every type τ (so that, effectively, the quantifiers were higher order properties; the properties of *having at least one term in its extension* and of *having every term in its extension* respectively). However, for the exploratory purposes of this paper I will omit any discussion of quantification.

The only additional constants worth mentioning in L are those used to express identity and definition. For every types τ_1, τ_2 there is a term $=$ of type $(\tau_1 \to (\tau_1 \to t))$, with the intended interpretation that $\ulcorner A^{\tau_1} =^{(\tau_1 \to (\tau_1 \to t)} B^{\tau_1}\urcorner$ means that A is identical to B,[20] and a term Def of type $(\tau_1 \to (\tau_2 \to t))$ with the intended interpretation that $\ulcorner Def^{(\tau_1 \to (\tau_2 \to t))}(A^{\tau_1}, B^{\tau_2})\urcorner$ means that A is, by definition, B. While I assume that terms must be of the same type in order to be identical, I do not assume that terms must be of the same type in order to be defined in terms of one another. Indeed, it is possible to make the stronger claim that, according to the ensuing theory, there are *never* terms of the same type such that one is defined in terms of the other—and this will hold the key to proving the irreflexivity of definition.[21]

Within L, it is possible to formally represent the aforementioned proxy for the structured proposition Fa—that is, the relation that only F stands in to a—as the bihaecceity:

$$\lambda X^{(e \to t)}.\lambda x^e.(X = F \wedge x = a)$$

Of course, there is nothing special about the proposition that Fa in particular; there are proxies for the structured propositions that Gb and Hc as well. It is thus desirable to provide a function that generates these proxies. This can be accomplished with the following:

$$\delta_1 := \lambda X^{(e \to t)}.\lambda x^e.\lambda Y^{(e \to t)}.\lambda y^e.(X = Y \wedge x = y).$$

The δ_1 function takes pairs of properties and objects as its input, and has—as its output—the relation that the input property stands in to the input object. This function can itself be generalized so that it provides proxies for terms of arbitrary type. For types τ_1, τ_2, there is a function:

$$\delta := \lambda X^{(\tau_1 \to \tau_2)}.\lambda x^{\tau_1}.\lambda Y^{(\tau_1 \to \tau_2)}.\lambda.y^{\tau_1}.(X = Y \wedge x = y)$$

[20] Within this paper, I shift freely between infix and prefix notation.

[21] In the following, I occasionally omit the types of the terms involved. This occurs either if the types are contextually evident—or else if the formulas are to be interpreted as schemata with applications in every type.

The δ function is to be interpreted as a schematic placeholder for functions of the various types: ones that generates proxies for terms of type τ_2. (So, for example, there are strictly different functions for proxies of monadic properties, binary relations, etc.). Although δ is more general than δ_1—which only generates proxies for terms of type t (indeed, merely a subset of those terms)—it still only applies to terms that stand in a particular functional relation to one another. That is, the inputs of δ must be two terms such that the latter is the functional input of the former. It will be necessary to have a function that is more general still: one that not only has outputs for terms of arbitrary type, but whose inputs are terms of arbitrary type as well. For types τ_1, τ_2, there is a function:[22]

$$\gamma := \lambda X^{\tau_1}.x^{\tau_2}.\lambda Y^{\tau_1}.y^{\tau_2}.(X = Y \wedge x = y)$$

With the γ function in place, the proxy for the proposition that Fa may be represented as:[23]

$$\gamma(F^{e \to t}, a^e)$$

It is useful to simplify this notation still further. The result of the γ function as applied to arbitrary inputs α, β may be represented as:

$$[\alpha, \beta]$$

In particular, the proxy for the proposition that Fa is to be represented as:

$$[F, a]$$

Because the δ function is a special case of the γ function, we can represent the outputs of δ with the [ ] notation—but cannot assume that terms within brackets stand in a particular functional relation to one another without additional information about their types.

It is also valuable to define recovery functions that take proxies as their input and have, as their output, the term it is a proxy for. We may (schematically) represent these functions as:[24]

[22] As with δ, γ should strictly be interpreted as a schema with instances in the various types.

[23] Here, as elsewhere, I assume that identity is preserved through $\beta\eta$ conversion—so that $\lambda X^{(e \to t)}.\lambda x^e.\lambda Y^{(e \to t)}.\lambda.y^e.(X = Y \wedge x = y)(F, a)$ is identical to $\lambda X^{(e \to t)}.\lambda x^e.(X = F \wedge x = a)$.

[24] This description of *Rec* might naturally be thought of as providing a restriction on the recovery function, rather than its complete definition. It specifies that $< \delta(\alpha, \beta), \alpha(\beta) >$ is in the extension of *Rec*—but not what else is in this function's extension. The intended interpretation is that *Rec* is undefined for any input that is not an instance of δ. We might attempt to denote this schema with 'the smallest family of functions that have $< \delta(\alpha, \beta), \alpha(\beta) >$ in their extension'—but this may fail to identify a unique relation depending on how fine-grained relations are. That is, if propositions are so fine-grained that there are distinct (yet logically equivalent) functions, several distinct functions might satisfy this description. Fortunately, this complication is irrelevant for our purposes. On coarse-grained accounts (which hold that only one relation validates 'the family of functions that satisfies this description and is undefined for any other input') we may identify *Rec* with this relation. On more fine-grained accounts, we can appeal to an arbitrary relation

$$Rec(\delta(\alpha, \beta)) = \alpha(\beta)$$

It may be more accurate to say that the γ function generates *immediate* proxies—and that, correspondingly, the *Rec* function generates the terms that they are immediate proxies for. However, it is also desirable to also represent chains of proxies: to have a function that is sensitive not only to the immediate syntactic structure of its inputs, but the mediate syntactic structure as well. For, just as it is impossible to recover property F from proposition Fa, so too it is impossible to recover property F from the proxy $[\neg, Fa]$ (though it remains possible to recover the $\neg$ operator and proposition Fa from this proxy). In the present context, the obvious way to describe this internal structure is via recursion. Let us define relations of decomposition, and say that one term may be decomposed into another. Decomposition is the smallest family of relations that satisfies the following:[25]

1. If, $\alpha = Rec(\delta(\beta, \psi))$, then $Dec(\alpha, [\beta, \psi])$
2. If $Dec(\alpha, [\beta, \psi])$ and $Dec(\beta, [\eta, \epsilon])$ then $Dec(\alpha, [[\eta, \epsilon], \psi])$
3. If $Dec(\alpha, [\beta, \psi])$ and $Dec(\psi, [\eta, \epsilon])$ then $Dec(\alpha, [\beta, [\eta, \epsilon]])$

There are two logically important attributes of decomposition as it has been defined:[26]

i)	Irreflexivity	There is no α such that $Dec(\alpha, \alpha)$.
ii)	Identity	If $Dec(\alpha, \eta)$ and $Dec(\beta, \eta)$, then $\alpha = \beta$.

Irreflexivity guarantees that nothing is a decomposition of itself—while *Identity* ensures that each decomposition is a decomposition of a unique term. But although each decomposition is a decomposition of a unique term, a term might have any number of decompositions; $\neg Fa$ may be decomposed as:

$$[\neg, Fa]$$

or, alternatively, as:[27]

that satisfies this description. My aim is to provide a logical restriction on definition that resolves the puzzles mentioned at the outset. Any function that satisfies this description performs this work.

Note, also, that the *Rec* function takes applications of the δ function—rather than the γ function—as its inputs, so we may indeed assume that the latter term is the functional input of the former.

[25] The reason I say 'family of relations' rather than 'relation' is that there are different decomposition relations for every type. The term '*Dec*' as it occurs within '$Dec(\neg Fa, [\neg, Fa])$' is of type $(t \to (((t \to t) \to (t \to t)) \to t))$, while the term '*Dec*' as it occurs within '$Dec(\neg Fa, [\neg, [F, a]])$' is of type $(t \to (((t \to t) \to (((e \to t) \to (e \to t)) \to t) \to t))$. Because these decompositions are of different types, they express different relations. As with the introduction of *Rec*, fine-grained accounts may modify the description of *Dec* to account for the fact that distinct relations satisfy this description.

[26] For proofs of these attributes, see appendix.

[27] This example underscores the need for the γ function. Had we operated only with the δ function, the expression $[\neg, [F, a]]$ would be ungrammatical, as the [ ] notation would require that the second term—$[F, a]$—be a functional input of the former term—$\neg$, which it isn't. By shifting to the γ function, we may

$$[\neg, [F, a]]$$

That is, $\neg Fa$ might be decomposed as the relation between negation and the proposition Fa, or, alternatively, as the relation between negation and the relation between F and a. At one end of the spectrum, decomposition is sensitive only to the outermost syntactic structure of a term, while—at the other end of the spectrum—a decomposition reveals the entirety of a term's syntactic structure. The latter of these is of particular interest, so it is valuable to define notation for that instance of decomposition. This notation need not be taken as primitive; it may be introduced recursively based upon the notation already to hand (defining it only for terms involving grammatically correct sequences two or more constants):

$\lceil AB \rceil = [A, B]$	for constants $A^{\tau_1 \to \tau_2}, B^{\tau_1}$
$\lceil A\beta \rceil = [A, \lceil \beta \rceil]$	for constant $A^{\tau_1 \to \tau_2}$ and non-constant β
$\lceil \alpha B \rceil = [\lceil \alpha \rceil, B]$	for non-constant α and constant B^{τ_1}
$\lceil \alpha\beta \rceil = [\lceil \alpha \rceil, \lceil \beta \rceil]$	for non-constants α, β

As it has been defined, there is a connection between the $\lceil\ \rceil$ function and decomposition; that is to say, if $\lceil \alpha \rceil = \beta$, then $Dec(\alpha, \beta)$.[28] This notation represents the final proxy for the structured proposition that $\neg Fa$ as:

$$\lceil \neg Fa \rceil = [\neg, \lceil Fa \rceil] = [\neg, [F, a]]$$

Note that the $\lceil\ \rceil$ notation is such that the types of these terms depend not merely upon the types of the terms occurring within its scope, but upon their syntactic structure as well; while $\lceil \neg\neg Fa \rceil$ strongly resembles $\lceil Fa \rceil$ the terms are of entirely different types. While the former denotes the relation that $\neg$ stands in to the relation that $\neg$ stands in to the relation that F stands in to a, the latter merely denotes the relation that F stands in to a.

Just as δ is a function that generates immediate proxies, so too Rec is a function that merely recovers the immediate terms that they are proxies for. While $Rec([\neg, Fa]) = \neg Fa$, this function is not defined in a manner that recovers $\neg Fa$ from $[\neg, [F, a]]$. We may also introduce a function that takes any proxy as its input and has—as its output—the coarse-grained term it is a proxy for, as the following:

$$\lfloor \alpha \rfloor = \beta \text{ iff } Dec(\beta, \alpha)$$ [29]

These formalisms suffice for the purposes of this paper. As is doubtlessly already clear, proxies make some of the same distinctions that structured propositions had been intended

express this type of proxy grammatically.

[28] For proof, see appendix.

[29] Given Identity (which is proven in the appendix), we can be confident that $\lfloor\ \rfloor$ is a function—i.e., that each proxy is a proxy for a unique term.

to make; for any α and β that differ in their syntactic structure, $\lceil\alpha\rceil$ is distinct from $\lceil\beta\rceil$. But we can be relatively confident that the theory of proxies is consistent as the theory of structured propositions is not. There are no resources within this language that surpass the descriptive power of traditional higher-order languages with λ abstraction and terms for identity. Rather, we have merely introduced shorthands to identify particular relations within this existing framework. This framework itself has been proven to be consistent—so the risk that these proxies engender contradiction is minimal.[30]

It is a matter of some interest why the theory of proxies is consistent, while the theory of structured propositions is not. What prevents us from constructing an analogous argument, and raising problems for the cardinality of proxies in just the manner we did for structured propositions?

The theory of proxies has resources that the theory of structured propositions lacks. On the structured view, propositions perform two theoretical roles: they are the bearers of truth and falsity, and they distinguish between any terms that differ in their syntactic structure. But on the proxy view, these tasks are divided; propositions are either true or false, while proxies make the fine-grained distinctions. One way to frame the upshot of Russell-Myhill is that the cardinality of the distinction-makers is higher than the cardinality of the truth-bearers. For the theory of structured propositions, this had the untenable consequence that the cardinality of propositions is higher than itself.[31] But the analogous result—that the cardinality of proxies is higher than the cardinality of propositions—is no problem at all.

Nevertheless, the ghost of Russell-Myhill impacts the expressive power of this theory. The problem depended upon the ability to assign every collection of propositions to a unique proposition—i.e., it was possible to assign each collection of propositions to the proposition asserting that every element of that collection is true. In the limiting case, there is a proposition asserting (albeit falsely) that all propositions are true. Proxies perform some of the theoretical functions of structured propositions. In order to generate an analogous problem for proxies, we would need the ability to make a claim about all proxies of arbitrary type. And this is something that the present theory can*not* do. Language L cannot

[30]The reason I say that we can be 'relatively confident' in the consistency of this language and that there is a 'minimal risk' of contradiction is that—as per Gödel's second incompleteness theorem—the systems used to prove the consistency of this language are more expressively powerful (and thus more likely to lead to contradiction) than language L itself. Because it is always possible to prove the consistency of inconsistent systems in this way, it not at all obvious how much these theorems ought to increase our confidence in consistency. But I note that—in this particular case—the system used to prove the consistency of the λ calculus is weaker than ZF set theory. So, if the system I use is ultimately inconsistent, then ZF set theory is inconsistent as well. This set theory forms the foundation of a vast number of branches of modern mathematics; if it is inconsistent, then all of these branches must be revamped from the ground up. Suffice it to say that no one has yet uncovered an inconsistency in this type of language and that—if there is one—the problems that that would generate would *far* exceed this paper.

[31]I suspect that the reason Fritz (Forthcoming) found an analogous puzzle for grounding is that it represents another area where propositions make fine-grained distinctions (in that they are individuated hyperintensionally in order for necessarily coextensive propositions to ground one another) and are truth-evaluable.

consistently have sufficiently many propositions that we can assign a unique proposition to each collection of proxies. And while it is possible to generate propositions that make claims about all proxies of type τ (for an arbitrary τ), it is impossible to generalize and thereby make a claim about proxies of any type. This is the cost of consistency—it is a price that I pay.

Definition by Proxy

At the outset, I claimed that definition gives rise to a number of puzzles—and suggested that decomposition holds the key to their resolution. Of course, decomposition itself has no interesting implications for definition; we require some principle linking the two. For the remainder of this paper, I will explore—and to some extent defend—the following:[32]

> DEFINITION BY PROXY (DBP)
> If $Def(\alpha, \beta)$ then $Dec(\alpha, \beta)$

If α is, by definition, β, then β is a decomposition of α. In order to adequately understand what this principle claims, it may help to briefly discuss what it *doesn't* claim. DBP could be strengthened in various ways. The conditional might, firstly, be substituted for a biconditional—so that $Def(\alpha, \beta)$ holds just in case $Dec(\alpha, \beta)$ holds. While DBP states that all definitions entail their corresponding decompositions, this stronger principle also states that all decompositions are definitions. And so, while one can consistently hold that DBP is true and $Def(Fa, [F, a])$ is false, this is not so for this stronger principle. The fact that Fa can be decomposed into $[F, a]$ ensures that the proposition is defined in terms of this proxy. Because DBP is given in conditional (rather than biconditional) form, it thus provides necessary conditions for definition—but not sufficient conditions.

The biconditional could itself be strengthened to constitute a definition of definition. That is, one may hold not only that 'definition' and 'decomposition' are coextensive, but also that definition is itself defined in terms of the relation of decomposition. On this view, $Def(Def(\alpha, \beta), [Dec, [\alpha, \beta]])$; for α to be, by definition β is, by definition, for α to be decomposed into β.[33] In some respects, this view may be considered to be hylomorphic—where definition is given in terms of matter and form. For a given property, proposition, or relation, the 'matter' of its definition is determined by the constants within our language L that figure in its definition and its form is determined by its logical form. For example,

[32]The reason I say 'to some extent' is that what I immediately aim to demonstrate is that this principle resolves the puzzles discussed at the outset of this paper. I do not contrast it with other conceptions of definition—nor do I claim that it satisfies any other theoretical requirements that definition ought to satisfy. Nevertheless, I take it that these problems are intractable enough that the ability to resolve them counts substantially in its favor.

[33]As previously mentioned in this paper, the fact that this can be expressed unparadoxically in this framework is an independent motivation to state a theory of definition in a typed, higher-order language like L.

if $Def(Fa, [F, a])$ then the matter of the definition of the proposition that Fa is given by the constants F and a, and the form is determined by the logical types of F $(e \to t)$ and a (e).

Each of these strengthened principles is independently worthy of consideration—but I will say no more about them here. It is my aim to resolve the puzzles mentioned at the outset of this paper and, to that end, DBP will suffice.

The Granularity of Definition

One puzzle concerned the granularity of definition. It was unclear how definition could make the fine-grained distinctions that metaphysicians typically take it to make. Quite plausibly, 'To be a triangle is, by definition, to be a polygon with three angles' is true while 'To be a triangle is, by definition, to be a polygon with three sides' is false—despite the fact that, necessarily, any object bearing one property also bears the other. If this is so, then an account of definition must be capable of distinguishing 'to be a polygon with three angles' from 'to be a polygon with three sides'—an extremely fine-grained distinction.

DBP is capable of making this distinction. If terms α, β differ in their syntactic structure, then $\lceil \alpha \rceil$ denotes a different proxy than $\lceil \beta \rceil$ denotes. For example, $\lceil p \wedge q \rceil$ differs in denotation from $\lceil q \wedge p \rceil$ (on the assumption that p and q are distinct); while the first denotes the relation that p stands in to the relation that $\wedge$ stands in to q, the second denotes the relation that q stands in to the relation that $\wedge$ stands in to p. And so, if definition entails decomposition it may be that $Def(p \wedge q, \lceil p \wedge q \rceil)$ is true while $Def(p \wedge q, \lceil q \wedge p \rceil)$ is false; the conjunction is defined in terms of one proxy but not the other. The proxies are distinct, so there is no obstacle to only one figuring in the definition of the conjunction.

The distinction between 'To be a polygon with three angles' and 'To be a polygon with three sides' is conceptually analogous, but the details are slightly more cumbersome due to the presence of 'three.' On a broadly Fregean approach, 'three' refers to the higher-order property of *is a property with three objects within its extension.* Technicalities aside, the proxy which relates this property to 'is a side' differs from the proxy that relates this to 'is an angle.' And, for this reason, it may be that 'To be a triangle is, by definition, to be a polygon with three angles' is true while 'To be a triangle is, by definition, to be a polygon with three sides' is false. DBP is thus capable of making extraordinarily fine-grained distinctions—indeed, as finely grained as the metaphysician could possibly require.

The Logic of Definition

A second puzzle concerned the logical structure of definition. There were three principles governing this logic that were each desirable, but that appeared to be incompatible:

Identification Hypothesis	If F is, by definition, G then there is an important sense in which F is the same as G.
Leibniz's Law	Terms that denote the same thing may be substituted for one another.
Irreflexivity	There are no reflexive definitions.

DBP allows for the consistent embrace of these principles. It entails that *Identification* and *Irreflexivity* are true—and does so in a manner that does not violate *Leibniz's Law*.

The easiest of these to establish is *Irreflexivity*. Armed with the formalisms we have developed, we can represent this as $\neg\exists X.Def(X,X)$. Given DBP, if there were a reflexive definition, then there would be a reflexive decomposition. That is, because definitions entail their corresponding decompositions, if there were an instance of $Def(F,F)$, then there would be an instance of $Dec(F,F)$. But we have already proven that there are no reflexive decompositions: no instances of $Dec(F,F)$.[34] Therefore, definition is irreflexive.

It is somewhat less clear how this accommodates *Identification*—in part because there is a reading of this principle on which it is false. If we interpret *Identification* as the claim that terms are literally identical to that which they are defined as, then this account falsifies this principle. There is no instance of $Def(\alpha,\beta)\wedge\alpha=\beta$. Indeed, if $Def(\alpha,\beta)$ then '$\alpha=\beta$' is not even *grammatical* in our language. For α to be, by definition, β, it must be the case that the terms are of different types—yet identity is introduced as a predicate that takes two terms of the same type.

But the claim that $Def(\alpha,\beta)$ entails that α is identical to β is only one way to interpret *Identification*. There can be (and, indeed, are) other important connections between definition and identity that vindicate this principle. In particular, DBP validates the following:[35]

$$Def(\alpha,\lceil\beta\rceil)\rightarrow\alpha=\beta$$

For an arbitrary term α, if α is, by definition, the final proxy for β, then α is identical to β. For example, if Fa is, by definition, $\lceil Gb\rceil$, then $Fa=Gb$.

This connection between definition and identity may seem inordinately restrictive: as not general enough to truly validate the *Identification Hypothesis*. After all, it does not apply to all definitions. DBP does not require terms to be defined by final proxies—some may be defined by proxies of intermediate granularity as well. For example, it may be that $\neg Fa$ is, by definition, $[\neg,Fa]$. The former connection between definition and identity says nothing about this case, so it may seem that the link between definition and identity is inadequate.

But there is another connection between definition and identity: one that applies not only to final proxies—but to them all:[36]

[34] For proof, see appendix.

[35] For proof, see appendix.

[36] For proof, see appendix.

$$Def(\alpha, \beta) \rightarrow \alpha = \lfloor \beta \rfloor$$

Therefore, the *Identification Hypothesis* is true.

Because DBP validates both *Irreflexivity* and *Identification*, it may seem certain to violate *Leibniz's Law*. And, indeed, it has implications that initially appear to ensure opacity. It may be that $Fa = Gb$ and $Def(Fa, \lceil Fa \rceil)$ are true while $Def(Fa, \lceil Gb \rceil)$ is false. For this reason, some might reasonably suspect that 'Fa' cannot be substituted for 'Gb' in every context—despite the fact that the two denote the same proposition.

This is no violation of *Leibniz's Law*. *Leibniz's Law* asserts that terms that denote the same thing may be substituted for one another *salve veritate*. But when the term 'Fa' appears within the $\lceil\ \rceil$ notation, it does not actually refer to the proposition Fa—rather, it refers to the relation that F stands in to a. And although Fa is identical to Gb, $\lceil Fa \rceil$ is not identical to $\lceil Gb \rceil$ if F is not identical to G. Because the terms $\lceil Fa \rceil$ and $\lceil Gb \rceil$ have different referents, *Leibniz's Law* is inapplicable.

To alleviate this concern, it may help to draw an analogy to quotation. Although 'Hesperus' and 'Phosphorus' both refer to Venus, the terms may not be substituted for one another when appearing within quotation marks. I.e., while Hesperus is Phosphorus and 'Hesperus' begins with the letter H, *Leibniz's Law* does not entail that 'Phosphorus' does as well. Within quotation, the terms refer to the words themselves rather than to their ordinary referents—and identity of referents does not guarantee the identity of words that denote them. In this respect, the $\lceil\ \rceil$ notation functions like quotation; a term has a different referent within this notation than it has out of it. While 'Fa' typically refers to the proposition that Fa, when it appears within $\lceil\ \rceil$ it refers to the relation F stands in to a. If we liked, we could eliminate the $\lceil\ \rceil$ notation from our language entirely, and represent proxies solely with $[\ ]$. Once this is done, the previous example becomes $Fa = Gb$ and $Def(Fa, [F, a])$ are true while $Def(Fa, [G, b])$ is false—which is no violation of *Leibniz's Law*.

The upshot is this: DBP entails that *Irreflexivity* and the *Identification Hypothesis* are true. And while it does not itself guarantee the truth of *Leibniz's Law*, the apparent conflict between the two is illusory; it thus provides no reason to reject *Leibniz's Law* beyond the puzzles it already faced. The ability to reconcile three seemingly incompatible principles counts in its favor.

The Paradox of Analysis

The final puzzle was the paradox of analysis—a puzzle that threatened the prospect of any substantive definition. A minimal requirement on substantiveness is that definitions are informative: there must be information within the content of analysis that is absent from the term being defined. But this difference in informational content appears to ensure that the putative definition is false. How could it be that the content and object are one and the same if they differ informationally? We need not bother constructing counterexamples to the claim that knowledge is, by definition, justified true belief. The fact that 'justified true belief' differs in meaning from 'knowledge' suffices to show that the putative analysis is false.

DBP wears its resolution on its sleeve. There is an important sense in which all definitions are substantive. Information is present in the content of analysis in that properties that figure in the definition can be extracted from this content. In contrast, the object being analyzed lacks this information—for the properties cannot be extracted from the object. If Fa is, by definition, $[F, a]$, then F is a term which can be extracted from the content of analysis that cannot be extracted from the object. And so, this definition reveals that F figures within this definition of this proposition—something which cannot be determined from Fa alone.

For the sake of concreteness, let us suppose that to be a brother is, by definition, to be a male sibling. In this case, *being a sibling* is a property that can be extracted from the content of analysis that cannot be extracted from the object being analyzed. Due to this difference in extractable information, the definition is substantive. However, this informational difference does nothing to undermine the *Identification Hypothesis*, which remains provably true.

Understanding why this informational difference is compatible with *Identification* requires carefully keeping track of the different terms. The term with additional information is, strictly, the proxy for *being a male sibling* (something along the lines of the relation that conjunction stands in to the relation that *being male* stands in to *being a sibling*)—as this is the sort of term from which *being a sibling* can be extracted. The property of *being a brother* is not identical to this proxy—for the simple reason that it isn't a proxy of any kind. It is, rather, a property of objects. Nevertheless, *being a brother* is identical to that which this proxy is a proxy for, and this accounts for the connection between definition and identity. *Identification* does not entail that $Def(Fa, [G, b]) \rightarrow Fa = [G, b]$, but it does entail that $Def(Fa, [G, b]) \rightarrow Fa = Gb$.[37] For this reason, there is an informational difference between the object and content of definitions that nevertheless entail identity.

This resolution of the paradox is not metalinguistic; the information contained within a definition is not merely information regarding the meanings of words. If someone learns that 'To be a square is, by definition, to be an equilateral rectangle,' they need not merely learn facts about the terms 'square' and 'equilateral rectangle.' Rather, they may learn facts regarding *being equilateral* and *being a rectangle*—in particular, that these properties figure in the definition of *being a square*—for that is information within the content of analysis which cannot be extracted from the object. This distinguishes this resolution from others in the literature, some of which do identify the information of substantive definitions metalinguistically.[38]

[37]Strictly, the expression '$Def(Fa, [G, b]) \rightarrow Fa = [G, b]$' is ungrammatical in L, as terms for identity were introduced only for terms of the same type, while Fa and $[G, b]$ are of different types—but I hope that this nevertheless conveys what I intend to express.

[38]A particularly notable example of this type of resolution occurs in Church (1954). This proposal is

An Objection—And Reply

There is an objection to the proposed resolution to the paradox of analysis that warrants a response.[39] The worry is that the paradox is merely an instance of the Frege/Mates puzzles of identity and synonymy. As such, however it is to be resolved ought to generate resolutions to these puzzles as well. But Definition by Proxy has no interesting implications for Frege/Mates, and so its resolution is unsatisfactory.

It is worth considering the objection in some detail. Frege's puzzle (at least in its initial formulation) concerns the distinction between trivial and substantive identity claims.[40] 'Hesperus is Hesperus' expresses a triviality: something knowable *a priori* (say those who countenance *a priori* knowledge), and something which is a mere instance of the universal truth that all objects are self-identical.[41] 'Hesperus is Phosphorus,' in contrast, expresses something substantive. It requires an enormous amount of astronomical evidence to verify that it is true, and its apparent logical form—that $x = y$—has many false instances. But if Hesperus really is Phosphorus, how could it be that 'Hesperus is Hesperus' differs in meaning from 'Hesperus is Phosphorus'?

Mates' puzzle is more general than Frege's. Mates noted that synonymous expressions do not appear to be substitutable for one another in every context. Let there be two synonymous—yet syntactically distinct—expressions D and D'. If synonymous expressions were substitutable, then 'Everyone who believes that D believes that D' would entail 'Everyone who believes that D believes that D''. It does not. To use an example alluded to in the epigraph, it may be that 'Function f is continuous at p' is synonymous with 'Given $\epsilon > 0$ there exists $\delta > 0$ such that if $|p - x| < \delta$ then $|f(p) - f(x)| < \epsilon$'—but these sentences surely cannot be substituted for one another in all belief ascriptions. But if synonymous expressions cannot be substituted for one another in every context, then they do not mean the same thing—which undermines the contention that they are synonymous.

Discussions of these puzzles are too extensive to canvas in any satisfactory way at present. They have rightly earned their place at the heart of the analytic tradition. But although potential resolutions are many and varied, there is one point of universal consensus: *something* must be done. There exists a resolution to these puzzles: one that explains

roundly (and, to my mind, quite rightly) critiqued for being metalinguistic in nature in Bealer (1982).

[39]My thanks to Peter Fritz for raising this objection.

[40]A correlative puzzle concerns the failure of *Leibniz's Law* in opaque contexts. Although Hesperus is identical to Phosphorus, and although Hesperus bears the property *being believed by the Babylonians to appear in the evening sky*, Phosphorus does not. As such, it seems that 'Hesperus' cannot be substituted by 'Phosphorus' in every context—and so *Leibniz's Law* is false.

[41]There are numerous iterations of this claim which could be discussed. Perhaps some believe that 'Hesperus is Hesperus' is only knowable *a posteriori*, because this sentence is true only if Hesperus exists, and it is not knowable *a priori* that Hesperus exists. Such philosophers may prefer to amend the present examples to 'Hesperus is Hesperus if Hesperus exists' and 'Hesperus is Phosphorus if Hesperus and Phosphorus exist' to arrive a difference in *a priori* knowledge. This discussion could doubtless be extended—as could the discussion of the distinction between *a priori* and *a posteriori* truths—but these discussions would take us farther afield than I care to venture.

the existence of substantive identity claims.[42] Given the *Identification Hypothesis*, it is not unreasonable to suspect that this resolution (whatever it may be) will have some bearing on accounts of definition. Perhaps the reason definitions are substantive is that their corresponding identity claims are substantive, and so the paradox of analysis can be neatly folded into Frege/Mates.[43]

This suggestion can be packaged into an argument against Definition by Proxy in the following way:

1. If a resolution to the paradox of analysis is adequate, then it resolves Frege/Mates as well.
2. DBP does not resolve Frege/Mates.
3. Therefore, DBP is not an adequate resolution to the paradox of analysis.

There are two premises in this argument—and, correspondingly, two places it may be resisted.

One possibility—which is all I intend to commit to at present—is to reject premise 1: the claim that if a resolution to the paradox of analysis is adequate, then it resolves Frege/Mates as well. Regardless of whether my proposed account of definition is to be accepted, it is at least a *potential* path toward resolving the paradox. Precisely since there is a potential resolution to the paradox of analysis that is not obviously a resolution to the Frege/Mates puzzles, the paradox is distinct from Frege/Mates. And because the paradox differs from Frege/Mates, we need not insist that a resolution to one possess the resources to resolve the other.

The other possibility is more interesting—but a fully satisfactory defense of it requires more space than I presently have. As such, what follows is merely a sketch of a proposal that is worthy of a paper in its own right. This possibility is to reject premise 2: the claim that DBP does not resolve Frege/Mates.[44] This rejection, of course, amounts to the claim that DBP *does* resolve Frege/Mates—and the explanation of how that resolution proceeds.

The resolution of Mates' puzzle is somewhat easier to grasp than the resolution of Frege's. Proxy-theory allows for there to be distinct proxies for the same sentence. Perhaps D expresses the same proposition as sentence D' (and, for this reason, the sentences are synonymous) while $\lceil D \rceil$ is distinct from $\lceil D' \rceil$ due to their different syntactic structures. Perhaps the apparent failures of substitution occur when the terms refer to proxies rather than to their ordinary referents.[45] That is to say, the reason why 'Everyone who believes

[42]Some (e.g., Braun (1988, 2002); Saul (2007)) argue that ordinary speakers are systematically mistaken about the truth-values of substantive identity claims—so 'Lois Lane believes that Superman can fly' entails that 'Lois Lane believes that Clark Kent can fly.' This may not count as allowing for substantive identity, but it is reasonable to expect these views to similarly translate to discussions of substantive definition.

[43]I myself suggested that this is so in Elgin (2020).

[44]My thanks to Michael Della Rocca for this suggestion.

[45]In some respects, this diagnosis resembles Frege's original suggestion that the term 'a' as it figures within '$a = b$' refers to the name of the object, rather than its ordinary referent. That is, he briefly held that

that D believes that D' does not entail that 'Everyone who believes that D believes that D'' is that the terms 'D' and 'D'' actually refer to '$\lceil D \rceil$' and '$\lceil D' \rceil$' within these sentences. When the terms occur under belief ascriptions, they refer to proxies for sentences, rather than to the sentences themselves. Because these proxies are distinct, they cannot be substituted for one another—but this does nothing to undermine the claim that D and D' are synonymous.[46]

We can make a similar move for Frege's puzzle. There are proxies for entities in the same way that there are proxies for properties and propositions; we may consider $\delta(A^{(t \to e)}, B^t)$ to be a proxy for terms of type e.[47] Moreover, just as there are numerous proxies for the same proposition, so too there are numerous proxies for the same entity. Perhaps apparent substitution fails in cases where names denote proxies for entities, rather than entities themselves. That is, perhaps the reason '$a = b$' differs in significance from '$a = a$' is that the first asserts that $\lceil a \rceil$ is a proxy for the same entity as $\lceil b \rceil$, while the second asserts that $\lceil a \rceil$ is a proxy for the same entity as itself.[48]

The upshot is this: the objection to DBP depended upon the claim that it does not resolve Frege/Mates. However, proxy-theory offers one path to resolving these puzzles. Those who find these resolutions tempting need not be persuaded by this objection.

Conclusion

The case is more-or-less complete. This account resolves three otherwise intractable puzzles: it resolves the paradox of analysis by providing information within the content of analysis that is absent from the object; it is the first account that can consistently embrace the *Identification Hypothesis, Leibniz's Law* and *Irreflexivity*—each of which is independently attractive; and it allows for metaphysicians to make extraordinarily fine-grained distinctions while avoiding problems that typically plague fine-grained accounts. I do not take these resolutions to constitute a conclusive argument in any robust sense. Philosophy rarely (if ever) meets that high bar. But it is enough, I think, to motivate a conception of definition

'$a = b$' expresses the proposition that 'a' and 'b' denote the same object. This suggestion is widely viewed as unsatisfactory—so I do not wish to place inordinate weight on this potential response. Nevertheless, I note that my proposal, unlike Frege's is *not* metalinguistic. Although 'D' changes its referents in certain contexts, it need not refer to the term 'D' in these cases.

[46]Some might maintain that Mates' puzzle can arise for atomic sentences as well; it may be that 'lawyer' and 'attorney' cannot be substituted for one another in every context, despite being synonymous. Because these terms do not have different syntactic structures, this resolution of Mates' puzzle presumably does not apply to this case. Those tempted by this resolution generally need find another explanation for these sorts of cases.

[47]This is by design. Often, higher-order languages are constructed so that $(t \to e)$ is not a type. I.e., it is held that e is a type and t is a type and for any types τ_1 and $\tau_2 \neq e$, $(\tau_1 \to \tau_2)$ is a type. In contrast, language L allows for $(t \to e)$ to be a type—and so terms of this type may figure in proxies for entities.

[48]We could, in an analogous manner, account for the apparent failure of the substitution of identicals in belief ascriptions in the obvious way.

by proxy.

References

Bacon, Andrew. 2019. "Substitution Structures." *Journal of Philosophical Logic* 48:1017–75.

Bacon, Andrew. Forthcoming. "A Theory of Structured Propositions." *The Philosophical Review* .

Bacon, Andrew and Jeffrey Russell. 2019. "The Logic of Opacity." *Philosophy and Phenomenological Research* 99(1):81–114.

Bealer, George. 1982. *Quality and Concept*. Oxford University Press.

Braun, David. 1988. "Understanding Belief Reports." *Philosophical Review* 107:555–95.

Braun, David. 2002. "Cognitive Significance, Attitude Ascriptions, and Ways of Believing Propositions." *Philosophical Studies* 108:65–81.

Caie, Michael, Jeremy Goodman and Harvey Lederman. 2020. "Classical Opacity." *Philosophy and Phenomenological Research* 101(3):524–66.

Church, Alonzo. 1954. "Intensional Isomorphism and the Identity of Belief." *Philosophical Studies* 5(5):65–73.

Correia, Fabrice. 2017. "Real Definitions." *Philosophical Issues* 27(1):52–73.

Correia, Fabrice and Alexander Skiles. 2019. "Grounding, Essence and Identity." *Philosophy and Phenomenological Research* 3:642–70.

Dorr, Cian. 2016. "To be F is to be G." *Philosophical Perspectives* 30(1):39–134.

Elgin, Samuel. 2020. "The Semantic Foundations of Philosophical Analysis." *The Review of Symbolic Logic* .

Elgin, Samuel. Forthcoming. "The Semantic Foundations of Philosophical Analysis." *The Review of Symbolic Logic* .

Fine, Kit. 1994. "Essence and Modality." *Philosophical Perspectives* 8:1–16.

Fine, Kit. 1995. "Ontological Dependence." *Proceedings of the Aristotelian Society* 95:269–90.

Fritz, Peter. 2021. "Structure by Proxy with an Application to Grounding." *Synthese* 198:6045–63.

Fritz, Peter. Forthcoming. "Ground and Grain." *Philosophy and Phenomenological Research* .

Myhill, John. 1958. "Problems Arising in the Formalization of Intensional Logic." *Logique et Analyse* 1:78–83.

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

Rosen, Gideon. 2015. "Real Definition." *Analytic Philosophy* 56(3):189–205.

Russell, Bertrand. 1903. *The Principles of Mathematics*. Cambridge University Press.

Saul, Jennifer. 2007. *Simple Sentences, Substitution, and Intuitions*. Oxford University Press.

Appendix

This appendix consists of proofs referenced throughout the paper.

Irreflexivity: There is no α such that $Dec(\alpha, \alpha)$

It may seem obvious (or, at least, relatively apparent) that decomposition is irreflexive given the way it is defined. Its base-case is determined by the recovery function; here, a term of type τ_2 is decomposed into a relation between terms of types $(\tau_1 \rightarrow \tau_2)$ and τ_1. Because a term is of a different type than the term it is decomposed into, the base case of decomposition is irreflexive. And the further we decompose a term (by appeal to the inductive steps), the further we ascend in the hierarchy of types—so the inductive instances of decomposition are irreflexive as well.

In order to codify this ascent, let us define a function *Rank* from the types to the natural numbers—such that $Rank(t) = 1$ and $Rank(e) = 1$ and, for types τ_1 and τ_2, $Rank(\tau_1 \rightarrow \tau_2) = Rank(\tau_1) + Rank(\tau_2)$. Trivially every type has the same rank as itself, so—to prove irreflexivity—it suffices to show that if $Dec(A^{\tau_1}, B^{\tau_2})$, then $Rank(\tau_1) \neq Rank(\tau_2)$. In particular, it will be shown that if $Dec(A^{\tau_1}, B^{\tau_2})$, then $Rank(\tau_1) < Rank(\tau_2)$. Let:

$Rank(\tau_1) = n$	$Rank(\tau_2) = m$	$Rank(\tau_3) = o$
$Rank(\tau_4) = p$	$Rank(\tau_5) = q$	

The proof proceeds by induction on *Dec*.

Base Case:

Suppose $\alpha^{\tau_2} = Rec(\delta(\beta^{(\tau_1 \rightarrow \tau_2)}, \psi^{\tau_1}))$.

Given the definition of *Rec*, $\alpha^{\tau_2} = \beta^{(\tau_1 \rightarrow \tau_2)}(\psi^{\tau_1})$.

The type of $[\beta, \psi]$ is $((\tau_1 \rightarrow \tau_2) \rightarrow (\tau_1 \rightarrow t))$.

$Rank(((\tau_1 \rightarrow \tau_2) \rightarrow (\tau_1 \rightarrow t))) = 2Rank(\tau_1) + Rank(\tau_2) + 1 = 2n + m + 1 > m$.

First Inductive Step:

Suppose $Dec(\alpha^{\tau_1}, [\beta^{\tau_2}, \psi^{\tau_3}])$ and $Dec(\beta^{\tau_2}, [\eta^{\tau_4}, \epsilon^{\tau_5}])$ such that $Rank(\tau_1) < Rank(\tau_2 \rightarrow (\tau_3 \rightarrow t))$ and $Rank(\tau_2) < Rank(\tau_4 \rightarrow (\tau_5 \rightarrow t))$.

Therefore, $n < m + o + 1$ and $m < p + q + 1$.

The type of $[[\eta, \epsilon], \psi]$ is $(((\tau_4 \rightarrow (\tau_5 \rightarrow t)) \rightarrow \tau_3) \rightarrow t)$.

The rank of this type is $p + q + o + 2$.

$p + q + o + 2 > m + o + 1 > n$.

Therefore, $Rank(\tau_1) < Rank(((\tau_4 \rightarrow (\tau_5 \rightarrow t)) \rightarrow \tau_3) \rightarrow t)$.

Second Inductive Step:

This can be demonstrated in an analogous manner to the first inductive step.

Therefore, Decomposition is irreflexive.

Identity: If $Dec(\alpha, \eta)$ and $Dec(\beta, \eta)$, then $\alpha = \beta$.

The proof proceeds by induction on Dec.

Base Case:

Trivially, if $\alpha = Rec(\delta(\phi, \psi))$ and $\beta = Rec(\delta(\phi, \psi))$ then $Dec(\alpha, [\phi, \psi])$ and $Dec(\beta, [\phi, \psi])$ and $\alpha = \beta$.

Inductive Step:

We will prove the first and second inductive steps simultaneously.

Suppose that ϕ, ψ have the inductive property—i.e., that for any ν, ζ, $Dec(\nu, \phi) \wedge Dec(\zeta, \phi) \rightarrow \nu = \zeta$ and $Dec(\nu, \psi) \wedge Dec(\zeta, \psi) \rightarrow \nu = \zeta$.

Suppose $Dec(\alpha, [\phi, \psi])$

Suppose $Dec(\beta, [\phi, \psi])$

Because ϕ and ψ have the inductive property, there must exist a unique ω, ϵ (which may or may not be identical to ϕ and ψ) that lack the [ ] notation such that $Dec(\alpha, [\omega, \epsilon])$ and $Dec(\beta, [\omega, \epsilon])$. That is, if ϕ lacks the [ ] notation then we may let $\omega = \phi$. If ϕ has the [ ] notation then it is a proxy for some term that lacks the [ ] notation. Given the inductive hypothesis, this term is unique—which we denote with ω. Just so for ψ and ϵ.

Because ω and ϵ lack the [ ] notation, it must be that $\alpha = Rec(\delta(\omega, \epsilon)) = \beta$.

Therefore, $Dec(\alpha, \eta) \wedge Dec(\beta, \eta) \rightarrow \alpha = \beta$.

The Link Between $\lceil\ \rceil$ and Decomposition: $\lceil \alpha \rceil = \beta \rightarrow Dec(\alpha, \beta)$

The proof proceeds by induction on $\lceil\ \rceil$:

Base Case:

Suppose that $A^{\tau_1 \rightarrow \tau_2}$ and B^{τ_1} are constants.

Then $\lceil AB \rceil = [A, B]$.

From the definition of Dec, it follows that $Dec(AB, [A, B])$.

First Inductive Step:

Select an $A\beta$ such that A is a constant and β is not a constant such that if $\lceil \beta \rceil = \eta$ then $Dec(\beta, \eta)$. Show that if $\lceil A\beta \rceil = \psi$ then $Dec(A\beta, \psi)$.

Select an arbitrary η such that $\lceil \beta \rceil = \eta$.

Recall that $\lceil\ \rceil$ is only defined over constants in function application position.

Therefore, it must be that β is a functional input of A.

From the definition of Dec, we have $Dec(A\beta, [A, \beta])$.

From the definition of Dec, it follows that $Dec(A\beta, [A, \eta])$.

$\lceil A\beta \rceil = [A, \lceil \beta \rceil] = [A, \eta]$.

Therefore, if $\lceil A\beta \rceil = \psi$ then $Dec(A\beta, \psi)$

Second and Third Inductive Steps:

These inductive steps are demonstrated in an analogous manner to the first inductive step.

Therefore, if $\lceil \alpha \rceil = \beta$ then $Dec(\alpha, \beta)$.

First Identification Hypothesis: $Def(\alpha, \lceil \beta \rceil) \rightarrow \alpha = \beta$

$Def(\alpha, \lceil \beta \rceil) \rightarrow \alpha = \beta$

Suppose $Def(\alpha, \lceil \beta \rceil)$. Given DBP, it follows that $Dec(\alpha, \lceil \beta \rceil)$.

Given a previous proof, it follows that $Dec(\beta, \lceil \beta \rceil)$.

We have $Dec(\alpha, \lceil \beta \rceil)$ and $Dec(\beta, \lceil \beta \rceil)$. Given Identity (from a previous proof), it follows that $\alpha = \beta$.

Second Identification Hypothesis: $Def(\alpha, \beta) \rightarrow \alpha = \lfloor \beta \rfloor$**:**

Suppose that $Def(\alpha, \beta)$.

Given DBP, we have that $Dec(\alpha, \beta)$.

Given the definition of $\lfloor \ \rfloor$ it follows that $\alpha = \lfloor \beta \rfloor$.