The Opacity of Definition

Samuel Z. Elgin

Abstract

Many metaphysicians maintain that real definition is both irreflexive and admits of substitution. I demonstrate that these commitments are at odds. If definition is substitutable, then there are reflexive definitions; cases in which something is defined directly and exclusively in terms of itself. As a corollary, I demonstrate that three central claims in 'Real Definition' Rosen (2015) are mutually inconsistent. I close with a brief discussion of the implications this has for the logic of definition and for philosophical methodology.

The notion of real definition is experiencing a renaissance within the analytic tradition. That old Socratic question 'What is justice?' can be interpreted as a demand for the definition of justice; the search for an analysis of knowledge is sometimes described as the quest for the definition of knowledge; physicalism might be understood as the claim that everything is defined in purely physical terms.[1] It is reasonable to take the object of investigation in these cases to be something worldly (such as a property or a state) rather than something linguistic (such as a word or sentence). On this conception, the answers to some of the deepest and most intractable questions that philosophy poses are definitions.

If this is correct, then the nature of definition is of paramount import. The standards that putative accounts must meet hang in the balance. While a definition of the good may determine whether a given state of affairs is a good one, a definition of definition determines what makes it the case that the good is defined as it is. And so, it is unsurprising that philosophers have attempted to provide reductive analyses of definition itself.[2] There is presently no consensus on what this definition of definition is. However, it is reasonable to expect a final theory to be accompanied by a logic of definition. With such a logic

[1] I have no doubt that some object to this characterization of physicalism due to the threat of multiple realizability. It may be that properties like *being in pain* or *being a heart* are defined functionally, rather than in terms of a particular physical configuration. Elsewhere, I argue that multiple realizability poses no such threat on the grounds that these interpretations are logically equivalent to ones well-suited to accommodate properties that can be multiply realized (see Elgin (Forthcoming)). However, I presently make no assumption that this the best—or even a reasonable—interpretation of physicalism. I merely assume that it is a way that one could interpret physicalism.

[2] Two examples I discuss in depth are Rosen (2015); Correia (2017). Others include Fine (2015); Horvath (2018).

at hand, we could potentially derive many answers from a few—quickly expanding our philosophical knowledge with ease.

The subject of this paper is this logic of definition. In particular, I am concerned with two logical features that are widely accepted, yet mutually incompatible. The first of these is *substitutability*. Some maintain that definition admits of substitution.[3] For example, if water is defined in terms of its atomic constituents, and its atomic constituents are defined in terms of their subatomic parts, it is reasonable to expect that water is ultimately defined in terms of its subatomic parts. Within the definition of water, the thought goes, water's atomic constituents can be substituted for their own definitions.

If this is so, it is presumably not peculiar to water. That is, if it holds at all, then it holds quite generally. Whenever one thing is defined in terms of something else, the latter may be replaced by its own definition within the definition of the former. So, if the property of *being a bachelor* is, by definition, the property of *being an unmarried male* and the property of *being unmarried* is, by definition, the property of *lacking a marriage*, then the property of *being a bachelor* is, by definition, the property of *being a male who lacks a marriage.* And if {2} is, by definition, the set containing only the number 2 and the number 2 is, by definition, the successor to the number 1, then {2} is, by definition, the set containing only the successor to the number 1.

A substitution principle is no idle tool—it may well perform substantial theoretical work. For example, if a physicalist were to demonstrate that biological properties are purely defined in terms of chemical properties, and that chemical properties are purely defined in terms of physical properties, she might employ a substitution principle to conclude that biological properties are defined purely in terms of physical properties. Substitution may thus be brought to the aid of physicalism. Conversely, if substitution principles fail, then demonstrating that all biological properties are chemically defined and that all chemical properties are physically defined would not guarantee that biological properties are physically defined. Something more would be required.

For what it's worth, I know of few metaphysicians who explicitly deny that definition is substitutable. Those who do not endorse substitution principles often take no stand whatsoever; rather, they discuss (and often endorse) the closely related phenomenon of transitivity.[4] Transitivity is strictly weaker than substitution—it allows one to infer from the fact that A is, by definition, B and the fact that B is, by definition, C that A is, by definition, C—but it does not allow one to 'dive into' B and replace some terms with others. If water is by definition the chemical compound H_2O, a transitivity principle would not itself allow the substitution of 'hydrogen' for its definition within the definition of water. Rather, it could only be employed if there were a definition of 'the chemical compound H_2O' as a whole. Nevertheless, I suspect that many who endorse transitivity would also subscribe to substitution principles if pressed.

[3] Substitution principles are presented against various theoretical backdrops. Three who can reasonably be read as endorsing substitution principles in this sort of context include Rosen (2015); Dorr (2016); Fine (2015).

[4] For example, see Correia (2017); Horvath (2018).

I deny that definition is substitutable. To be clear, I do not claim that there are no cases in which substitution succeeds. Rather, I claim that general substitution principles conflict with another logical attribute of definition: *irreflexivity*. As I will argue, even modest substitution principles entail that there are reflexive definitions—cases in which something is defined directly and exclusively in terms of itself. I reject reflexive definitions, and suspect that I am in good company in doing so.[5] Strange as the literature on personal identity may be, it has never been suggested that Socrates' definition is Socrates himself. And while some have argued that knowledge is primitive, I know of no one who has suggested that knowledge is, by definition, knowledge. However, I have no new argument against reflexive definitions at present, so all that I take myself to demonstrate is a conditional: *if* definition admits of substitution, *then* there are reflexive definitions. I am personally more averse to reflexive definitions than wed to substitution principles, so I take this to constitute a reason to abandon definitional substitution. Those who do not object to reflexive definitions, however, may retain substitution principles.

This might seem unsurprising. Suppose we were to license the following kind of inference, without yet taking a stand on what definition consists of:

A is, by definition, B

C is, by definition, D

$\therefore$ (A is, by definition, B)$^{[C/D]}$

That is to say, if A is, by definition B and C is, by definition D, then any replacement of C with D within 'A is, by definition, B' is valid. This principle straightforwardly accommodates the previous examples.[6] If {2} is, by definition, the set containing only the number 2 and the number 2 is, by definition, the successor to the number 1, this principle licenses the inference to: {2} is, by definition, the set containing only the successor to the number 1.

However, this principle quickly generates reflexive definitions. In fact, it generates reflexive definitions so transparently that I know of no one who actually endorses it. Suppose, for example, that the property of *being a vixen* is, by definition, the property of *being a female fox*. This principle can be employed to derive that the property of *being a female fox* is, by definition, the property of *being a female fox*—a reflexive definition. And

[5] As with substitution, metaphysicians express their commitment to irreflexivity against various theoretical backdrops. However, those who can reasonably be read as endorsing definitional irreflexivity include, e.g., Rosen (2015); Correia (2013, 2017); Fine (1995, 2015); Horvath (2018). An instance of a philosopher who endorses the related phenomenon of asymmetry is Wildman (2017). I suspect that many detractors do not take definition to be reflexive, but rather maintain that identity ought to perform the theoretical work often attributed to definition. Such philosophers (may) include Rayo (2013); Dorr (2016); Correia and Skiles (Forthcoming).

[6] For some, slight alterations might be required to preserve grammaticality. I assume that these alterations preserve the spirit behind the principle.

quite generally, by allowing the same example to witness the first two conditions, it is possible to derive reflexive definitions.

Perhaps for this reason, metaphysicians often appeal to more restricted principles. Reflexivity arose, they might reasonably suspect, because the substitution principle licensed substitution within the definiendum—or object of analysis. If a principle were restricted to the definiens—or content of analysis—then reflexivity could be avoided.

It is my aim to demonstrate that this is false; this restricted substitution principle *also* gives rise to reflexive definitions. To that end, I present two prominent conceptions of definition. The first, championed by Rosen (2015), is that definition is a relation between relations and structured complexes. For example, it might be a relation that holds between the property of *being a father* and the complex of being a male parent. The second, advanced by Correia (2017), is that real definition is a kind of generalized identity. For example, 'To be even is to be a natural number divisible by 2 without remainder' might express a definition. On each account, substitution principles give rise to reflexive definitions. I then argue that the tension between substitution and irreflexivity is generalizable: every account of definition has the very same result. As such, I am not primarily concerned with whether the accounts provided by Rosen and Correia are correct. For my purposes, they merely serve as foils—as useful steps toward a general conclusion.

I proceed as follows. In section 2, I present the position that definition is a relation between properties and complexes, before arguing that substitution principles give rise to reflexive definitions on this view. *En route*, I demonstrate that three central claims in 'Real Definition' Rosen (2015) are logically inconsistent. In section 3, I present the position that real definitions are a type of generalized identity, before showing that substitution principles give rise to reflexive definitions on this view. In section 4, I generalize the arguments from sections 2 and 3 to argue that substitution principles generate reflexive definition on any account of definition. In section 5, I present a proof that definition is substitutable (strange as it may seem to do so). I do so for two reasons: firstly, because substitution principles are owed their due; it is important to not merely consider what can be said against them, but also what can be said in their favor. Secondly (and, in my mind, much more importantly), I present this proof to highlight the costs of abandoning substitution principles. If substitution fails, one or more of the premises that the proof relies upon must also be false. I conclude in section 6 by highlighting some implications this has for philosophical methodology.

I do not canvass everything that might reasonably be called a 'substitution principle' here. Perhaps some avoid reflexive definitions. Nevertheless, I take it that the principles I discuss are rightly called 'substitution principles,' and, I suspect, are what many have in mind when discussing definitional substitution.

The Relational Conception

Some philosophers have come to suspect an intimate connection between the notions of real definition and ground.[7] If {Socrates} is defined in terms of set membership and Socrates, it seems no accident that '{Socrates} exists' is made true by something about set membership and Socrates. So, perhaps the identity of a thing—what it is that makes it the thing that it is—is connected to that which makes the proposition that it exists true. And if the notions of definition and ground are so closely related, it is also natural to suspect that one is defined in terms of the other.

To the best of my knowledge, the first sustained development of this suspicion occurred in Rosen (2015), who provided a definition of definition in terms of ground. Rosen maintains that definition is a relation holding between properties (or relations) and structured complexes. He provides no account of structured complexes, claiming that they are "built from worldly items in roughly the sense in which a sentence is built from words" (pg. 190). These complexes resemble Russellian propositions except for the presence of free variables corresponding to the adicity of the property or relation being defined. For example, if the property of *being a person* is, by definition, *being a rational animal*, then the definition relation obtains between the property of *being a person* and the complex of *being a rational animal*.

Problems may already arise. If these complexes so closely resemble structured propositions, it is reasonable to expect them to inherit any obstacles that structured propositions face. Structured propositions have experienced a sustained onslaught in recent years, primarily on logical grounds.[8] One of their central tenets is that if the proposition Fa is identical to the proposition Gb, then $a = b$ and $F = G$. 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 the property of *being a brother* is identical to the property of *being a male sibling.* This tenet is radically incompatible with an orthodox principle of higher-order logic: propositional identity is preserved through β-conversion. Consider, for example, the binary property F of *being the same height as*; two people stand in relation F just in case they are equally tall. β-equivalence entails that $\lambda x.Fxx(a) = \lambda x.Fxa(a)$. That is to say, 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—and that $\lambda x.Fxx = \lambda x.Fxa$—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.[9] Those who endorse structured propositions (and, I suspect, Rosen's related notion of structured complexes) are thus committed to revising higher-order logic. However, as previously mentioned, I am not primarily concerned with whether Rosen's account is ultimately correct, so I set these types of worries aside.

[7] See, e.g., Fine (2015).
[8] See, e.g., Dorr (2016); Goodman (2017); Caie, Goodman and Lederman (Forthcoming)
[9] To his credit, Rosen foresees (but largely sets aside) worries stemming from structured propositions.

In developing this account, Rosen adopts a factive notion of ground. The fact that a ball is both red and round might be grounded in the fact that it is red and the fact that it is round, and the fact that an act is morally right might be grounded in the fact that it maximizes expected utility. Ground is taken to be a primitive, non-causal type of metaphysical dependence. Various locutions express similar ideas; if fact A grounds fact B we might say 'B holds *because* A holds' or perhaps 'B holds *in virtue of* A's holding.' Although the notion of ground has recently come under fire, it remains a mainstay of contemporary metaphysics.[10]

Rosen adopts the standard (but not uncontroversial) assumptions that grounding is a many-one relation and a strict-partial ordering; i.e., that it is transitive, irreflexive and asymmetric.[11] Ground is symbolized with '$\leftarrow$,' so '$A \leftarrow B$' means that the fact that A is grounded in the fact that B. Importantly, $\leftarrow$ represents full, rather than partial ground. It may be that the fact that John is male partially grounds the fact that John is a bachelor, while the fact that John is male and the fact that John is unmarried collectively fully ground the fact that John is a bachelor. On this account full, rather than partial, ground is at issue.

The other notational convention Rosen adopts that surpasses the language of second-order logic is the indexed modal operator $\Box_t$, where 't' is a singular term. This operator first appears in Fine (1995) and is roughly translated as 'it is necessary in virtue of the identity of t.' So, for example, '$\Box_{Knowledge}$ knowledge is a mental state' means that it is necessary in virtue of the identity of knowledge that knowledge is a mental state, and '$\Box_{Better\text{-}Than}$ *Better-than* is transitive' means that it is necessary in virtue of the identity of the relation *better-than* that *better-than* is transitive.

Thus armed, Rosen defends the following account:

$$Def(F, \phi) \text{ iff } \Box_F \forall x((Fx \vee \phi x) \rightarrow (Fx \leftarrow \phi x))$$

This asserts that property F is defined in terms of complex ϕ just in case it is necessary in virtue of the identity of F that, for all objects, if that object is either F or ϕ, then it is F in virtue of being ϕ (or, rather, the fact that it is ϕ grounds the fact that it is F). For example, the property *being morally right* is, by definition, *maximizing utility* just in case it is necessary in virtue of the identity of *being morally right* that if an act is morally right or maximizes utility, then the fact that it maximizes utility grounds the fact that it is morally right. Although Rosen presents his account as a biconditional, he intends to provide something stronger: the very definition of definition. That is to say, definition is itself a relation, and it is defined in terms of the condition Rosen presents.

[10] For dissenters, see e.g., Della Rocca (2014); Wilson (2014). For responses, see, e.g., Schaffer (2016); Berker (2018)

[11] For challenges to the many-one conception, see, e.g., Dasgupta (2014). For challenges to transitivity, see, e.g., Schaffer (2012) (but for a response see Litland (2013)). For challenges to irreflexivity and asymmetry see, e.g., Jenkins (2011).

Notably, this account does not entail that a definiendum is identical to its definiens. Even if the property of *being morally right* is so defined, it may not be identical to the complex of *maximizing utility*. If the relevant complexes are not themselves properties or relations, an application of Leibniz's Law ensures that the definiendum is distinct from its definiens. After all, the definiendum bears the property *being a property or relation* while its definiens does not. In this sense, Rosen's account is not reductionist. Even if every chemical property were defined in terms of physical complexes, it may be that (at least some) chemical properties remain distinct from these physical complexes.

Rosen explicitly endorses definitional substitution, claiming:

> It should be possible to prove a principle that licenses arbitrary definitional expansion: If Def(F, ϕ) and Def (G, ψ) then Def($F, \phi^{\psi/G}$) where $\phi^{\psi/G}$ is the result of substituting ψ for G in ϕ...Any account of real definition should license the substitutions of definiens for definiendum in a ground to yield a further ground. (pg. 201).

Rosen thus maintains that definitional substitution is not only admissible, but ought to be provable.[12] His substitution principle is restricted to the definiens; it is admissible only within the content of analysis, not within the object of analysis. If the property of *being a square* is, by definition, *being an equilateral rectangle* and the property *being a rectangle* is, by definition, *being a four-sided polygon*, this principle licenses the inference to the claim that the property of *being a square* is, by definition, *being an equilateral four-sided polygon*. However, by disallowing definitional substitution within the definiendum, Rosen avoids the inference to the claim that the property of *being a vixen* is, by definition, the property of *being a vixen* in the manner previously discussed.

Rosen also maintains that definition is irreflexive, claiming:

> The ground-theoretic account of real definition has a number of appealing features...It explains why trivial definitions are excluded. Why are there no cases in which *Def(F, F)*? Because there are no cases in which $Fa \leftarrow Fa$, i.e., because grounding is irreflexive. (pg. 201)

Both ground and definition are irreflexive, Rosen maintains, and the irreflexivity of definition arises from the irreflexivity of ground. And so, Rosen and I disagree; while he endorses both definitional substitution and irreflexivity, I claim that one of the two must go.

This is as good a point as any to note that Rosen's commitments are logically inconsistent. In particular, the following three claims engender contradiction:

[12] However, Rosen provides no such proof—perhaps he has in mind something analogous to the proof I present in section 5.

a) Grounding is irreflexive.

b) Definition is substitutable.

c) $Def(F, \phi)$ iff $\Box_F \forall x((Fx \vee \phi x) \rightarrow (Fx \leftarrow \phi x))$

Drawing out this inconsistency quickly becomes technically cumbersome, so I help myself to the following terminological abbreviation:

$$\mu = \Box_F \forall x((Fx \vee \phi x) \rightarrow (Fx \leftarrow \phi x))$$

μ is thus shorthand for the content of Rosen's definition of definition. Nothing philosophically substantive turns on this shorthand; the inconsistency could straightforwardly be articulated without it. Nevertheless, it would quickly become technically unwieldy, so I help myself to the abbreviation. Rosen's definition of definition thus becomes:

$$Def\,(F,\phi) \text{ iff } \mu \tag{1}$$

Property F is, by definition ϕ just in case μ. Applying Rosen's account to itself (i.e., by taking definition itself to be the relation subject to definition) yields the following:

$$Def(Def, \mu) \text{ iff } \Box_{\text{Def}} \forall_{<F,\phi>}((Def(F, \phi) \vee \mu) \rightarrow (Def(F, \phi) \leftarrow \mu)) \tag{2}$$

That is to say, definition is defined in terms of μ just in case it is necessary in virtue of the identity of definition that, for all F and ϕ, if either F is defined in terms of ϕ or μ, then the fact that F is defined by ϕ is grounded in the fact that μ. This simply results from applying Rosen's account to itself; from taking it to be a definition of—rather than merely a necessary and sufficient condition for—definition. The substitution principle and (1) collectively license the substitution of occurrences of '*Def(F,* ϕ*)*' with 'μ' within the definiens of (2). One such application results in the following:

$$Def(Def, \mu) \text{ iff } \Box_{\text{Def}} \forall_{<F,\phi>}((Def(F, \phi) \vee \mu) \rightarrow (\mu \leftarrow \mu)) \tag{3}$$

(1) contains a reflexive grounding relation: $\mu \leftarrow \mu$. Therefore, there is at least one reflexive grounding relation: the fact that μ grounds the fact that μ. So Rosen's definition of definition and the substitution principle collectively contradict the claim that there are no reflexive grounding relations. Because Rosen is committed to the claims that definition is substitutable and irreflexive (and to his definition of definition), his commitments are mutually inconsistent. In order to retain his account, he must either abandon the substitutability of definition or the irreflexivity of ground.

Troubling as this inconsistency is, our present concern is with the irreflexivity of definition, rather than the irreflexivity of ground. Rosen crucially relies upon the irreflexivity of ground in arguing that definition is irreflexive; if he abandons the irreflexivity of ground in order to preserve the substitution principle, there is no guarantee that definition is irreflexive. But is there more? Is it possible to demonstrate that substitution principles engender reflexive definitions, rather than merely reflexive grounding relations?

Employing the substitution principle twice more yields the following:

$$\mathrm{Def}(\mathrm{Def}, \mu) \text{ iff } \Box_{\mu} \forall_{<F, \phi>}((\mu \vee \mu) \rightarrow (\mu \leftarrow \mu)) \tag{4}$$

This asserts that definition is defined in terms of μ just in case it is necessary in virtue of the identity of μ that for all ordered pairs $<F, \phi>$, if either μ or μ, then the fact that μ is grounded in the fact that μ. This, of course, is precisely what is required for μ to be defined in terms of itself, i.e.:

$$\mathrm{Def}(\mu, \mu) \text{ iff } \Box_{\mu} \forall_{<F, \phi>}((\mu \vee \mu) \rightarrow (\mu \leftarrow \mu)) \tag{5}$$

From (4), (5), and the symmetry and transitivity of the classical biconditional, we obtain:

$$\mathrm{Def}(\mathrm{Def}, \mu) \text{ iff } \mathrm{Def}(\mu, \mu) \tag{6}$$

Therefore, if definition is defined in terms of μ, then μ is defined directly and exclusively in terms of itself.[13] And so, on the relational conception of definition, substitution principles give rise to reflexive definitions.

The Generalized Identity Conception

Some may suspect that the tension between irreflexivity and substitution is peculiar to Rosen's account. After all, the inconsistency depended not only on the two logical principles, but also the content of Rosen's analysis. Various philosophers have objected to other aspects of this proposal. In addition to the worries about structured propositions that were previously discussed, Correia (2017) objects on the grounds that the account does not guarantee that a definiendum is identical to its definiens. Perhaps a more plausible conception of definition can retain both substitution and irreflexivity principles.

To that end, I set Rosen's account aside and consider a prominent alternative. This alternative concerns a targeted reading of sentences of the form 'To be *F* is to be *G*' which

[13] Because Rosen is also committed to the claim that there are no reflexive definitions, this is another point of inconsistency.

has received considerable attention in recent years and in which the 'is' closely resembles the 'is' of identity. Notable examples include 'To be a bachelor is to be an unmarried male' and 'To be just is to be such that each part of one's soul does its own proper work.' These sentences have borne multiple labels in the literature. Some refer to them as 'Just-is statements' (e.g., Rayo (2013); Linnebo (2014)), others prefer 'identifications' (e.g., Dorr (2016); Caie, Goodman and Lederman (Forthcoming)) while still others call them 'generalized identities' (e.g., Correia (2017); Correia and Skiles (Forthcoming)). Nothing significant turns on our preferred label, so long as the targeted reading itself is clear; I will refer to them as 'generalized identities' here.

Because the 'is' of generalized identities so closely resembles the 'is' of identity, some might suspect that these sentences to express identities of properties. That is to say, perhaps 'To be F is to be G' is true just in case the property of being F is identical to the property of being G. Several philosophers object to this interpretation.[14] Minimally, there is linguistic evidence that 'to be F' is not synonymous with 'the property F.' The sentence 'I hope to be an accomplished philosopher' is perfectly true, but 'I hope the property of being an accomplished philosopher' is not. Additionally, some desire accounts of generalized identities to be compatible with nominalism: the denial that abstract objects like properties exist. If 'To be F is to be G' were synonymous with 'The property F is identical to the property G,' nominalists would lack the resources to appeal to generalized identities from the outset. For these reasons, some maintain that generalized identities closely resemble, but are not strictly, identity claims.

Nevertheless, these sentences are standardly taken to share the logical and modal profile of identity. In particular, they are often assumed to be reflexive, transitive and symmetric and if they are true, then they are necessarily true and it is necessary that all and only Fs are Gs. Because these sentences are symmetric, they cannot be immediately identified with definitions (under the assumption that definition is asymmetric).[15] If to be a square is, by definition, to be an equilateral rectangle, it cannot be that to be an equilateral rectangle is, by definition, to be a square.

Correia (2017) argues that definitions are a kind of generalized identity; they are generalized identities that satisfy a further condition. So, the set of definitions is a subset of the set of generalized identities. Correia offers two ways of winnowing down the generalized identities into those that are definitions—one which employs the notion of ground and the other which employs Lewis's notion of relative naturalness.[16] The connection between these methods is independently worthy of investigation, but does not impact the issue of substitution, so I restrict my attention to his characterization in terms of ground. Correia's notion of ground differs slightly from Rosen's. In particular, he does not employ a factive notion of grounding. The kinds of things which stand in grounding relations are generics

[14] E.g., Dorr (2016); Correia (2017).

[15] For an early argument that these sentences cannot perform the theoretical work attributed to definition for this reason, see Cameron (2014).

[16] For the development of relative naturalness see, canonically, Lewis (1983).

(like *being F* and *being G*) rather than facts.[17] With this notion of ground at hand, Correia defends the following account:

To be F is$_{df}$ to be G iff:
a) To be F is$_{id}$ to be G[18]
b) Being G grounds being F

'To be even is to be divisible by 2 without remainder' expresses a definition just in case 'to be even is to be divisible by 2 without remainder' also expresses a generalized identity and being divisible by 2 without remainder grounds being even. On this account, a sentence expresses a reflexive definition just in case it both expresses a definition and the terms for F and G are identical: i.e., just in case it takes the form 'To be F is$_{df}$ to be F.' If 'To be morally right is to be morally right' expresses a definition, then it expresses a reflexive definition. A constraint against definitional reflexivity amounts to the claim that no sentences express reflexive definitions.

Some interpretive issues warrant attention before turning to the conflict between substitution and irreflexivity. It is unclear whether Correia intends his account to be a definition of definition or merely a generalized identity of definition. That is, he does not specify whether his account merely identifies the property of being a definition, or whether it provides the grounds of definition as well. Here, I interpret his account as a definition of definition. Recall that, for the present purposes, it serves as a foil—a mere step towards a general result. The most useful step for us to take is to treat this as a definition of definition. Further, unlike Rosen, Correia does not explicitly endorse definitional substitution (although he does embrace the related phenomenon of transitivity). This too is unimportant for our present purpose. My aim is not to demonstrate that Correia's claims are false, but rather to show how definitional substitution conflicts with irreflexivity on this account.

Let us assume the following restricted substitution principle:

$$\frac{\text{To be } F \text{ is}_{df} \text{ to be } G \qquad \text{To be } H \text{ is}_{df} \text{ to be } I}{\therefore \text{ To be } F \text{ is}_{df} \text{ to be } (G^{[H/I]})}$$

If 'To be hydrogen is$_{df}$ to be the element containing exactly one proton' and 'To be a proton is$_{df}$ to be the subatomic particle consisting of two up quarks and one down quark,' this principle entails that 'To be hydrogen is$_{df}$ to be the element containing exactly one subatomic particle consisting of two up quarks and one down quark.'

[17]More precisely, they are representations of generics, but this refinement is not needed here.
[18]Following Correia, I distinguish the reading of 'To be F is to be G' that resembles an identity from the reading that resembles a definition with the subscripts '*id*' and '*df*.'

I assume that it is permissible to conjoin Correia's two criterion into one: To be F is$_{df}$ to be G iff To be F is$_{id}$ to be $G \wedge$ being G grounds being F. As with Rosen's account, I employ terminological abbreviation in order to avoid excessive formalism:

$$\omega = \textit{To be F is}_{id} \textit{ to be G} \wedge \textit{being G grounds being F}$$

That is to say, ω is shorthand for the content of Correia's account. The account thus becomes:

$$\textit{To be definition is}_{df} \textit{ to be } \omega \tag{7}$$

Applying this account to itself results in:

$$\textit{To be definition is}_{df} \textit{ to be } \omega \textit{ iff to be definition is}_{id} \textit{ to be } \omega \wedge \textit{being } \omega \textit{ grounds being definition} \tag{8}$$

As before, this simply results from taking this account to be a definition of, rather than a necessary sufficient condition (or a generalized identity) for definition. The substitution principle and (7) collectively license replacing occurrences of 'definition' with ω in the definiens of (8). An application of this results in:

$$\textit{To be definition is}_{df} \textit{ to be } \omega \textit{ iff to be definition is}_{id} \textit{ to be } \omega \wedge \textit{being } \omega \textit{ grounds being } \omega \tag{9}$$

As with the relational conception, the substitution principle generates a reflexive grounding relation on Correia's proposal—being ω grounds being ω. Another application of substitution results in:[19]

$$\textit{To be definition is}_{df} \textit{ to be } \omega \textit{ iff to be } \omega \textit{ is}_{id} \textit{ to be } \omega \wedge \textit{being } \omega \textit{ grounds being } \omega \tag{10}$$

This is what is required for ω to be defined in terms of itself, i.e.:

$$\textit{To be } \omega \textit{ is}_{df} \textit{ to be } \omega \textit{ iff to be } \omega \textit{ is}_{id} \textit{ to be } \omega \wedge \textit{being } \omega \textit{ grounds being } \omega \tag{11}$$

From (10), (11), and the symmetry and transitivity of the classical biconditional, we thus obtain:

$$\textit{To be definition is}_{df} \textit{ to be } \omega \textit{ iff to be } \omega \textit{ is}_{df} \textit{ to be } \omega \tag{12}$$

[19] Although the substitution principle licenses this inference, it is not strictly necessary for the derivation of (10) from (9). Because generalized identities are reflexive, we already were committed to the claim that to be ω is$_{id}$ to be ω. However, the substitution principle was required to derive (9) from (8).

Therefore, if definition is defined in terms of ω, then ω is defined directly and exclusively in terms of itself. Definitional substitution thus gives rise to reflexive definitions on the generalized identity account. Correia may not endorse the substitution principle I mention—he does not explicitly embrace it here. He would do to disavow it, as it generates reflexive definitions.

The Universal Link

Substitution principles generate reflexive definitions on two prominent conceptions of definition. I maintain that this not peculiar to these two views: any definition of definition will have the very same result.

The first indication that this is so is that the same conflict occurs on both Rosen's and Correia's accounts. Both employ some notion of ground or other, but differ in many other respects. Rosen's notion of ground is factive, while Correia's is generic; Correia's account guarantees identity, while Rosen's does not. It would be surprising if the very same tension coincidentally arose on such different views. Another indication is that terminological abbreviations were possible. On each account, it was possible to derive the tension while treating the content of analysis as a point—ignoring any of the complexities within. That is to say, we were able to describe the entirety of these accounts with a single symbol, and derive reflexive definitions without considering the details of what that symbol represents. This indicates that these details are not responsible for the conflict between substitution and irreflexivity.

The general connection between substitution and reflexivity can be shown more directly. Select an arbitrary definition of definition: a theory that dictates how definition is itself defined. Let us denote the content of this account—whatever it may be—with τ; definition is, by definition, τ. Definition stands in some relation or other to τ, and it is in virtue of standing in this relation that it is defined in terms of τ. A substitution principle will allow replacing occurrences of 'definition' with occurrences of 'τ' within any definiens. In particular, when articulating the definiens of definition itself (i.e., when articulating the relation definition stands in to τ), a substitution principle will permit replacing 'definition' with 'τ.' Therefore, a substitution principle will guarantee that the relation definition stands in to τ is the same as the relation τ stands in to itself. Because standing in this relation suffices for definition to be defined in terms of τ, it also suffices for τ to be defined in terms of τ. And so, because definition is defined in terms of τ, a substitution principle will entail that τ is defined in terms of τ. Because the selection of τ was arbitrary, any definition of definition is such that substitution principles give rise to reflexive definitions.

The derivations for the relational and generalized identity accounts followed this precise template. The relational account stipulated that definition is defined in terms of μ, while the generalized identity account maintained that definition is defined in terms of ω.

Substitution principles guaranteed that the relation definition stands in to μ is the same as the relation μ stands in to μ (on the former), and that the relation definition stands in to ω is the same as the relation ω stands in to ω (on the latter). For this reason, they entailed that μ is defined in terms of μ and that ω is defined in terms of ω respectively. All accounts have this same result—if definition is substitutable, then there are reflexive definitions.

A Proof of Substitution

In this section I offer a proof that definition is universally substitutable. This proof was inspired by, and is somewhat analogous to, the proof of substitution of identicals recently presented in Caie, Goodman and Lederman (Forthcoming). I have no doubt this seems a strange—perhaps even bizarre—thing for me to include, given the former argument for rejecting substitution principles. My primary aim is to demonstrate that definitional substitution is not inferentially isolated. If we are to abandon substitution principles (as I believe we should), we must abandon other logical principles as well, as these principles entail that substitution succeeds.

Before presenting this proof, I briefly discuss the principles it relies upon. This discussion falls far short of a full-throated defense of these principles; they ought to be controversial. Numerous philosophical views are committed to the claim that some are false, and because I reject substitution, I myself must abandon at least one. Nevertheless, as the saying goes, one person's modus ponens is another's modus tollens. Some might reasonably interpret this discussion as an excuse to embrace reflexive definitions, rather than a reason to abandon the principles.

This proof occurs in a typed, higher-order language with λ-abstraction. There is one basic type e of entities, and for any $n \geqslant 0$ and types $\tau_1, \tau_2 ...., \tau_n$, $< \tau_1, \tau_2 ...., \tau_n >$ is a type. In this language, monadic first-order predicates can be identified with objects of type $< e >$, dyadic first-order predicates can be identified with objects of type $< e, e >$, etc.. Propositions are the limiting case of entities typed $<>$, the $\neg$-operator is of type $<<>>$, and the binary operators $\wedge, \vee, \rightarrow, \leftrightarrow$ are all of type $<<>, <>>$. In addition, this language is equipped with infinitely many variables of each type, and the corresponding λ-abstracts needed to bind them. There is also a symbol $=_{df,\,\tau}$ for each type τ, with the intended interpretation that '$\alpha_\tau =_{df\ \tau} \beta_{\ \tau}$' means that α is, by definition, β. In what follows, I often omit indications of type in the principles I appeal to. In these cases, I mean to indicate that these principles are schemas with applications in every type.

The principles I rely upon are the following:

Material Abstraction: $\phi =_{df} \psi \rightarrow \lambda x.\phi[x/a] =_{df} \lambda x.\psi[x/a]$

Lift-Application: $(a =_{df} b \wedge \phi =_{df} \psi) \rightarrow \lambda X.Xa(\phi) =_{df} \lambda X.Xb(\psi)$

Beta-Eta Equivalence: $\phi =_{df} \psi \rightarrow \varphi =_{df} \psi$ provided ϕ and φ are beta-eta equivalent.

$\phi =_{df} \psi \rightarrow \phi =_{df} \varphi$ provided ψ and φ are beta-eta equivalent.

Let us take these in turn. Material Abstraction provides a way of abstracting away from a particular bit of matter and, in so doing, altering the type of the objects being defined. If the proposition that John is a bachelor is, by definition, the proposition the John is an unmarried male, Material Abstraction entails that the property of *being a bachelor* is, by definition, the property of *being an unmarried male*. If the proposition that Megan is next to Abby is, by definition, the proposition that Megan is adjacent to Abby, Material Abstraction entails that the property of *being next to Abby* is, by definition, the property of *being adjacent to Abby*. The underlying thought behind this principle is that the particular bits of matter are not responsible for these definitions. If the proposition that Sarah is a sister is, by definition, the proposition that Sarah is a female sibling, this is not because of anything about Sarah. Rather, the proposition is defined as it is because of the definition of the property *being a sister*.

Some philosophical views are incompatible with Material Abstraction. A notable example is a kind of moral particularism.[20] Moral particularists deny that there are general right-making features that apply to all acts. Instead, they contend that the morality of acts must be determined in isolation. That which makes an act of giving to charity right might differ from that which makes an act of saving a drowning child right. A moral particularist might claim that the proposition that act a is morally right is, by definition, the proposition that act a has features F (if the features F are the right-making features of this particular act), while denying that the property of being morally right is, by definition, the property of having features F. This amounts to the rejection of Material Abstraction.

Additionally, there are difficult cases for Material Abstraction to contend with. Suppose that the proposition that the number 4 is even is, by definition, the proposition that the number 4 is divisible by 2 without remainder. Material Abstraction could be employed to show that the property of being even is, by definition, the property of being divisible by 2 without remainder, and this is plausible enough. However, suppose instead that the proposition that the number 2 is even is, by definition, the proposition that the number 2 is divisible by 2 without remainder. In this case, Material Abstraction could be employed to show that the property of being even is, by definition, the property of being divisible by oneself without remainder; Material Abstraction replaces all occurrences of '2' with the same variable bound by the λ-abstract. This is clearly false.

Turn to Lift-Application. This is the definitional analog of combining two principles appearing in Caie, Goodman and Lederman (Forthcoming)—Lift Congruence and Application Congruence:

Lift Congruence: $a =_{df} b \rightarrow (\lambda X.Xa) =_{df} (\lambda X.Xb)$

Application Congruence: $F =_{df} G \rightarrow Fa =_{df} Ga$

[20]For example, see Dancy (1983, 2004); Lance, Potrč and Strahovnik (2008).

These are plausible principles of identity (although Caie, Goodman and Lederman discuss the implications of dropping each at length), but seem much less plausible when applied to definition. For example, if the number 3 is, by definition, the successor to the number 2, Lift-Congruence entails that the proposition that 3 is odd is, by definition, the proposition that the successor of the number 2 is odd. If Socrates is, by definition, the person resulting from this sperm and that egg, Lift-Congruence entails that the fproposition that Socrates is human is, by definition, the proposition that the person resulting from this sperm and that egg is human.[21] Some might deny that these constitute definitions on the grounds that a definition of the proposition that the number 3 is odd ought to concern the definition of *being odd*, as well as the number 3 and that a definition of the proposition that Socrates is human ought to concern the definition of *being human* as well as Socrates.

This kind of worry is allayed if Lift-Congruence is immediately followed by a predicate and its own definition. Lift-Application entails that the definition of the proposition that 3 is odd concerns both the definition of the number 3 and the definition of the property *being odd*. And if Socrates is by definition, the result of this sperm and that egg and the property of *being human* is, by definition, the property of *being a rational animal*, Lift-Application entails that the proposition that Socrates is human is, by definition, the proposition that the result of this sperm and that egg is a rational animal.

Lift-Application resembles a commitment of structured propositions previously discussed: $Fa = Ga \rightarrow a = b \wedge F = G$. As mentioned before, this principle is incompatible with Beta-Eta Equivalence. For obvious reasons, it would be undesirable to employ two incompatible principles. However, Lift-Application is not the definitional analog of the commitment to structured propositions: it is its converse. To the best of my knowledge, it is perfectly compatible with Beta-Eta Equivalence.

Beta-Eta Equivalence is the definitional analogue of an orthodox principle of higher-order logic—that identity is preserved under beta-eta conversion.[22] This is most often defended indirectly; i.e., by dispelling potential counterexamples, rather than directly arguing in its favor. I have little to add to these defenses. Because Beta-Eta equivalence is incompatible with the existence of structured propositions, its primary detractors are those who endorse propositional structure.

However, I note one complication that arises in treating Beta-Eta Equivalence as a principle of definition, rather than identity. It entails that propositions are not defined in terms of Beta-conversion (assuming there are no reflexive definitions). For if $\lambda x.Fx(a) =_{df} Fa$, Beta-Eta Equivalence would entail $Fa =_{df} Fa$. Those who maintain that propositions *are* defined in terms of Beta-conversion ought to deny this principle (or embrace reflexivity). However, I see no reason to suspect that propositions are defined in terms of beta-conversion.

With these principles at hand, we can prove that definition is substitutable. The proof

[21]Strictly speaking, these examples only follow if Beta-Eta Equivalence is also assumed. In what follows, I assume it holds.

[22]See, e.g., Dorr (2016); Caie, Goodman and Lederman (Forthcoming).

proceeds as follows:

1. $a =_{df} b$ Premise
2. $\phi =_{df} \psi$ Premise
3. $\lambda x.\phi[x/a] =_{df} \lambda x.\psi[x/a]$ 2, Material Abstraction
4. $\lambda X.Xa(\lambda x.\phi[x/a]) =_{df} \lambda X.Xb(\lambda x.\psi[x/a])$ 1, 3 and Lift-Application
5. $\lambda x.\phi[x/a](a) =_{df} \lambda x.\psi[x/a](b)$ 4, Beta-Eta Equivalence
6. $\phi =_{df} \psi[b/a]$ 5, Beta-Eta Equivalence

Therefore, if Material Abstraction, Lift-Application and Beta-Eta Equivalence are all correct, then definition admits of substitution. As I have argued, definitional substitution engenders reflexive definitions. So, if these principles are correct, then there are reflexive definitions. Those like myself, who reject reflexive definitions, must abandon one or more of these principles. I leave the discussion of which principle ought to be abandoned for further investigation.

Conclusion

Restricted substitution principles are employed to retain the theoretical benefits of definitional substitution while avoiding the cost of reflexivity. The primary conclusion of this paper is that these principles fail to achieve the desired result. On any conception of definition, substitution principles generate reflexive definitions.

I close by mentioning some of the implications this has. One, already discussed at length, is that abandoning substitution requires abandoning other logical principles as well. A second implication is methodological. Unless philosophers countenance reflexive definitions, they ought not appeal to substitution principles when presenting philosophical arguments. For example, if philosophers were to argue that sets are defined in terms of their members and that people are defined in terms of their genetic makeup, they ought not immediately infer that {Socrates} is defined in terms of Socrates' genetic makeup. And if biological properties are defined in terms of chemical properties and chemical properties are defined in terms of physical properties, it does not immediately follow that biological properties are defined in terms of physical properties.

Another implication depends on other philosophical commitments. Fine (1995) has argued that ontological dependence ought to be understood in terms of definitional containment.[23] Loosely, entity A ontologically depends upon entity B just in case B is contained within the definition of A. The reason water ontologically depends upon hydrogen is that

[23]For objections to this proposal, see, e.g., Koslicki (2012); Wilson (Forthcoming).

hydrogen is contained within the definition of water. On this account, substitution principles guarantee that ontological dependence is transitive. Because *B* ontologically depends upon everything contained within its definition, and because that definition may be substituted within the definition of *A*, *A* ontologically depends upon all entities contained within the definition of *B*. Without a substitution principle, we have no guarantee that ontological dependence is transitive.

Yet another implication is that definition is linguistically opaque (if it is irreflexive). While transparent contexts are ones in which co-referential terms can be substituted salva-veritate, opaque contexts are ones in which they cannot. Many (presumably most) predicates are transparent. So long as 'Superman' denotes the same person as 'Clark Kent,' one can infer from 'Superman is 6′ 3″ tall' that 'Clark Kent is 6′ 3″ tall.' In contrast, 'believe' is a canonical opaque predicate. While 'Lois Lane believes that Clark Kent works for the Daily Planet' may be true, 'Lois Lane believes that Superman works for the Daily Planet' may be false. If there are no reflexive definitions, then 'definition' functions more like 'believes' than 'is 6′ 3″ tall,' precisely because substitution principles fail.

I suspect there are still further implications. In particular, I believe that the opacity of definition sheds light on the paradox of analysis, which concerns how a definition can be both true and informative if its definiens is identical to its definiendum. A systematic discussion of this implication is worthy of a paper in its own right, so I leave that topic for another time.

References

Berker, Selim. 2018. "The Unity of Grounding." *Mind* 127(507):729–777.

Caie, Michael, Jeremy Goodman and Harvey Lederman. Forthcoming. "Classical Opacity." *Philosophy and Phenomenological Research* .

Cameron, Ross. 2014. "On the Lack of Direction in Rayo's *The Construction of Logical Space*." *Inquiry* 59(4):382–97.

Correia, Fabrice. 2013. Metaphysical Grounds and Essence. In *Varieties of Dependence*, ed. E. Hoeltje, B. Schnieder and A. Steinberg. Philosophia Verlag pp. 271–91.

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

Correia, Fabrice and Alexander Skiles. Forthcoming. "Grounding, Essence and Identity." *Philosophy and Phenomenological Research* .

Dancy, Jonathan. 1983. "Ethical Particularism and Morally Relevant Properties." *Mind* 92:530–47.

Dancy, Jonathan. 2004. *Ethics without Principles*. Clarendon Press.

Dasgupta, Shamik. 2014. "On the Plurality of Grounds." *Philosopher's Imprint* 14.

Della Rocca, Michael. 2014. "Razing Structures to the Ground." *Analytic Philosophy* 55(3):276–94.

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

Elgin, Samuel. Forthcoming. "Physicalism and the Identity of Identity Theories.".

Fine, Kit. 1995. "Ontological Dependence." *Proceedings of the Aristotelian Society* 95:269–290.

Fine, Kit. 2015. "Unified Foundations for Essence and Ground." *Journal of the American Philosophical Association* 1(2):296–311.

Goodman, Jeremy. 2017. "Reality is Unstructured." *Analysis* 77(1).

Horvath, Joachim. 2018. "Philosophical Analysis, the Concept Grounding View." *Philosophy and Phenomenological Research* 97:724–50.

Jenkins, C. S. 2011. "Is Metaphysical Dependence Irreflexive?" *The Monist* 94(2):267–76.

Koslicki, Kathrin. 2012. Varieties of Ontological Dependence. In *Metaphysical Grounding, Understanding the Structure of Reality*, ed. Fabrice Correia and Benjamin Schnieder. Cambridge University Press.

Lance, Mark, Matjaž Potrč and Vojko Strahovnik. 2008. *Challenging Moral Particularism*. Routledge.

Lewis, David. 1983. "New Work for a Theory of Universals." *Australasian Journal of Philosophy* 61(4):343–77.

Linnebo, Øystein. 2014. "'Just is'-Statements as Generalized Identities." *Inquiry* 57(4):466–482.

Litland, Jon. 2013. "On Some Counterexamples to the Transitivity of Ground." *Essays in Philosophy* 14(1):19–32.

Rayo, Agustín. 2013. *The Construction of Logical Space*. OUP Oxford.

Rosen, Gideon. 2015. "Real Definition." *Analytic Philosophy* 56(3):189–209.

Schaffer, Jonathan. 2012. Grounding, Transitivity and Contrastivity. In *Metaphysical Grounding: Understanding the Structure of Reality*, ed. Benjamin Correia, Fabrice & Schnieder. Cambridge University Press pp. 122–38.

Schaffer, Jonathan. 2016. Ground Rules: Lessons from Wilson. In *Scientific Composition and Metaphysical Ground*, ed. Kenneth Aizawa and Carl Gillett. Springer Nature.

Wildman, Nathan. 2017. "Against the Reduction of Necessity to Essence." *Synthese* .

Wilson, Jessica. 2014. "No Work for a Theory of Grounding." *Inquiry* 57(5-6):535–79.

Wilson, Jessica. Forthcoming. Essence and Dependence. In *Metaphysics, Meaning and Modality*, ed. Mircea Dumitru. Oxford University Press.