Real Definition is Either Unsubstitutable or Circular

Samuel Z. Elgin

July Draft

The concept of real definition plays an integral role not only in metaphysics, but in nearly all branches of philosophy.[1] Some moral disputes can be understood as disagreements over the definition of the good, attempted amendments of a justified-true-belief analysis of knowledge as the search for a definition of knowledge, and (one type of) physicalism as the claim that everything is defined in purely physical terms. The nature of definition is thus of central philosophical import.

Some may suspect that definition is substitutable. If water is defined in terms of its atomic constituents, and its atomic constituents are defined in terms of their subatomic structure, it is reasonable to maintain that water can ultimately be defined in terms of its subatomic structure. Within the definition of water, the thought goes, we may substitute water's atomic constituents with their definition, *salva veritate*.

I maintain that this is false; real definition is not substitutable. It is not my claim that there are *no* cases in which substitution succeeds. Rather, it is my claim that even modest substitution principles entail definitional circularity. At present, I have no new argument against circular definitions, so all that I actually demonstrate in this paper is that either definitions are unsubstitutable or are circular. However, I myself am more averse to definitional circularity than I am wed to definitional substitution, so I take this to be a reason to abandon substitution principles.

I am unsure of how surprising this result will be. Definitions are hyperintensional—distinct, yet logically equivalent. As philosophers, we have observed many cases in which substitution principles fail for hyperintensional entities. Nevertheless, it remains valuable to flesh the relevant cases out in detail—to understand why it is that substitution fails in this particular case.

Let us become more precise. Let us say that some sentences *express* definitions. In particular, let us assume that the sentences that express definitions are a subset of those that express identities. While 'water is the chemical compound H_2O' plausibly expresses a definition, 'nitrogen is the most prevalent element in the Earth's atmosphere' does not.

[1] Real definition is the metaphysical analogue of nominal definition. While the nominal definition of 'water' concerns the definition of the term, the real definition is a definition of the chemical compound itself. While some dispute the distinction between real and nominal definition, I assume that it is coherent for the purposes of this paper. Here, I use 'definition' to refer to real definition unless otherwise specified.

Sentences that express definitions have three components: a term for the *definiendum*—or object of analysis, a term for the *definiens*—or content of the analysis, and an identity sign linking the two. In English, the definiendum typically appears on the left of the identity sign, while the definiens typically appears on the right. The examples that I use here all distinguish the definiendum from the definiens by that ordering.

Let us say that a sentence *expresses a circular definition* just in case the terms for the definiendum and definiens are identical—that is to say, if and only if the sentence takes the form $\ulcorner\phi =_{df} \phi\urcorner$. A constraint against circularity, then, amounts to the claim that no sentences express circular definitions.

Additionally, let us assume that there is at least one sentence that expresses a definition. We need not take any stand on what this sentence is, but grant that at least one exists. Quite trivially, if there were no sentences that express definitions then there would be no sentences that express circular definitions. However, I take the situation in which substitution and anti-circularity principles are both vacuously true to be uninteresting, and so I set that possibility aside for the purposes of this paper.

It is also helpful for us to be able to speak of a term's *partial content*. This is meant without the implications concerning analyticity that are typically associated with 'content'. Rather, a term ϕ is a part of the content of a term ψ just in case it literally appears within ψ. So, 'unmarried' is a part of the content of 'unmarried male', but is not a part of the content of 'bachelor'. Here, my use of 'partial content' refers to improper part, rather than to proper part, so every term is a part of its own content.

Perhaps the simplest substitution principle is the following:

Full Substitution Principle:
For any sentence s that expresses a definition, if there exist terms ϕ and ψ such that:
i) $\ulcorner\phi =_{df} \psi\urcorner$ expresses a definition
ii) ϕ is a part of the content of a term within s
then substituting an occurrence of ϕ for ψ within s results in a sentence that also expresses a definition.

This principle quickly generates sentences that expresses circular definitions. In fact, it generates definitional circularity *so* quickly that I have never seen anyone actually endorse it. The problem arises once we allow s to be the sentence $\ulcorner\phi =_{df} \psi\urcorner$. Assume, for example, that 'to be a vixen $=_{df}$ to be a female fox' expresses a definition. Nothing hinges on this particular sentence—the result holds for whichever example we select. The derivation of circularity is as follows:

1. 'to be a vixen $=_{df}$ to be a female fox' expresses a definition (stipulation)
2. 'to be a vixen' is a part of the content of 'to be a vixen' (from the definition of parthood)

3. 'to be a female fox $=_{df}$ to be a female fox' expresses a definition (from 1, 2 and Full Substitution Principle)
4. There exists a sentence that expresses a circular definition (from 3 and the definition of sentences that express circular definitions)

If Full Substitution were true, sentences that express circular definitions would be common. For every term ϕ that figures as the definiens in a sentence that expresses a definition, $\ulcorner\phi =_{df} \phi\urcorner$ expresses a definition.[2] Plausibly, this occurs because Full Substitution licenses substitution within the definiendum of a sentence. If substitution were restricted to the definiens, one might reasonably think, circularity could be avoided. Such a restricted principle is the following:

> **Restricted Substitution Principle:**
> For any sentence s that expresses a definition, if there are terms ϕ and ψ such that:
> i) $\ulcorner\phi =_{df} \psi\urcorner$ expresses a definition
> ii) ϕ is a part of the content of the definiens of s
> then substituting an occurrence of ϕ for ψ within the definiens of s results in a sentence that also expresses a definition.

Unfortunately, Restricted Substitution also engenders definitional circularity. Let us consider this circularity in its general form, before moving to a particular case. The real trick here is the ability to embed one definition within the definiens of another. Once this is done, circularity quickly results.

1. $\ulcorner\varphi =_{df} \pi\urcorner$ expresses a definition (stipulation, and our candidate sentence s)
2. $\ulcorner\phi =_{df} \psi\urcorner$ expresses a definition (stipulation)
3. $\ulcorner\phi =_{df} \psi\urcorner$ is a part of the content of π (stipulation)
4. $\ulcorner\phi =_{df} \psi\urcorner$ is a part of the content of the definiens of $\ulcorner\varphi =_{df} \pi\urcorner$ (from 3 and definition of definiens)
5. $\ulcorner\varphi =_{df} \pi'\urcorner$, such that π' differs from π only in that it contains $\ulcorner\psi =_{df} \psi\urcorner$ instead of $\ulcorner\phi =_{df} \psi\urcorner$ expresses a definition (from 2, 4 and Restricted Substitution Principle)

So φ is defined in terms of a sentence that expresses a definition, and is also defined in terms of a sentence that expresses a circular definition. However, in this case, there is a sentence that expresses a circular definition. A particular case illustrates the problem. Let

[2] This is not to say that the full substitution principle entails that all occurrences of $\ulcorner\phi =_{df} \phi\urcorner$ express definitions. If there is a term A that never figures as a definiens in a sentence that expresses a definition, it does not entail that '$A =_{df} A$' expresses a definition.

us suppose that there are propositions, and that they are defined in terms of their logical form, i.e. that

$$`p \wedge q =_{\mathit{df}} \wedge (p, q)\text{'}$$
$$`p \vee q =_{\mathit{df}} \vee (p, q)\text{'}$$
$$`\neg p =_{\mathit{df}} \neg(p)\text{'}$$

all express definitions. $p \wedge q$ is, by definition a conjunction—in particular, the conjunction of the proposition that p with the proposition that q. $p \vee q$ is, by definition, a disjunction—in particular the disjunction of the proposition that p with the proposition that q. $\neg p$ is, by definition, a negation—in particular the negation of the proposition that p.

I assume that there is a proposition asserting that $p \wedge q$ is defined in terms of its logical form. Let us denote this proposition as $[p \wedge q =_{\mathit{df}} \wedge (p, q)]$, and consider the conjunction of *that* proposition with proposition r.

1. '$r \wedge [p \wedge q =_{\mathit{df}} \wedge (p, q)] =_{\mathit{df}} \wedge (r, [p \wedge q =_{\mathit{df}} \wedge (p, q)])$' expresses a definition (from the definition of propositions in terms of logical form).
2. '$p \wedge q =_{\mathit{df}} \wedge (p, q)$' expresses a definition (from the definition of propositions in terms of logical form).
3. '$p \wedge q =_{\mathit{df}} \wedge (p, q)$' is a part of the content of the definiens of '$r \wedge [p \wedge q =_{\mathit{df}} \wedge (p, q)] =_{\mathit{df}} \wedge (r, [p \wedge q =_{\mathit{df}} \wedge (p, q)])$' (from the definition of definiens).
4. '$r \wedge [p \wedge q =_{\mathit{df}} \wedge (p, q)] =_{\mathit{df}} \wedge (r, [\wedge(p, q) =_{\mathit{df}} \wedge (p, q)])$' expresses a definition (from 2, 3 and Restricted Substitution Principle).

Note that the definiens of the sentence in 4 contains '$\wedge(p, q) =_{\mathit{df}} \wedge (p, q)$'. If this expresses a definition, then it expresses a circular definition. If we allow r to be any true proposition, then the definiendum in 4 denotes a proposition that is true, and, if there are no circular definitions, the definiens denotes a proposition that is false—the conjunction of a true proposition (r) with a false proposition ($[\wedge(p, q) =_{\mathit{df}} \wedge (p, q)]$). So, unless there are sentences that express circular definitions, sentence 4 does not even express an identity, much less a definition! Restricted substitution thus engenders sentences that entail circular definitions.

This example relies on a particular account of propositions. However, it should be clear that similar problems arise for any example in which one definition may be built within the definiens of another.

This circularity does not arise from the stipulation that sentences that express definitions are a subset of those that express identities. Accounts of definition not in terms of identity have circularity worries as well (though the ways these transpire differs slightly from the cases discussed above). The most prominent such account is provided by Gideon Rosen (**Reference**). I apologize for the lengthy and cumbersome formalisms that follow—I have been unable to uncover a more straightforward derivation of circularity than this.

Two components of Rosen's formulation exceed classical logic. The first is the '$\leftarrow$' symbol, which denotes the grounding relation. Grounding is often taken to be a relation of metaphysical explanation, however Rosen introduces as the in-virtue-of relation. Plausibly, the fact that two is prime holds in virtue of the fact that its only factors are one and itself, the fact that a molecule is water holds in virtue of the fact that it is the chemical compound H_2O, etc. On this understanding, the fact that 2 is prime is *grounded* in the fact that its only factors are one and itself, and the fact that a molecule is water is *grounded* in the fact that it is the chemical compound H_2O, etc. So, on the symbolism, the grounded appears on the left of the '$\leftarrow$' symbol, while the grounds appear on the right. So '$p \leftarrow q$' denotes the claim that the fact that q grounds the fact that p.

The second piece of symbolism is the $\Box_F$ operator. This symbolism is borrowed from Fine (**Reference**). This operator denotes necessity that arises from a particular source—in this case, from the identity of F. Plausibly, it is necessary in virtue of the identity of {Socrates} that {Socrates} is a set, but it is not necessary in virtue of the identity of {Socrates} that the law of excluded middle holds. So '$\Box_{\{\text{Socrates}\}}$ {Socrates} is a set' is true, while '$\Box_{\{\text{Socrates}\}}$ the law of excluded middle holds' is false.

Rosen's definition of definition is the following:

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

In some ways, this statement is obscure. Although Rosen stipulates at the outset that his aim is to define definition, and claims that 1 is a definition of definition, the actual content of 1 is stated in terms of necessary and sufficient conditions; F is, by definition ϕ if and only if... Nevertheless it strikes me as reasonable to treat this as his definition of definition; i.e. as his proposed definition of the definition relation. If we grant that second-order relations (like the definition relation) are defined analogously to first-order relations, this amounts to the following:

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

Let us grant, for the sake of argument, that substitution is permissible within a definiens. That is to say, the following holds:

Rosen Substitution Principle[3]
For any properties F and G such that
i) $Def(F, \phi)$
ii) $Def(G, \psi)$
iii) G occurs as a constituent of ϕ
G may be substituted by ψ within ϕ, *salva veritate*.

An application of this principle results in the following:

[3] With this name I do not mean to assert that Rosen has endorsed this principle—rather that it is a substitution principle as it applies to Rosen's account

3. $Def(Def, \Box_F \forall x((Fx \vee \phi x) \to (Fx \leftarrow \phi x)))$ iff
$\Box_{Def} \forall F((Def(F, \Box_F \forall x((Fx \vee \phi x) \to (Fx \leftarrow \phi x))) \vee \Box_F \forall x((Fx \vee \phi x) \to (Fx \leftarrow \phi x)))$
$\to (\Box_F \forall x((Fx \vee \phi x) \to (Fx \leftarrow \phi x)) \leftarrow \Box_F \forall x((Fx \vee \phi x) \to (Fx \leftarrow \phi x))))$.

3 contains a reflexive grounding relation—there is a fact that grounds itself. A loose, watered-down version of 3 asserts that Rosen's account of definition holds just in case it is necessary in virtue of the nature of definition that, for any property F, if it is the case the F is, by definition $\Box_F \forall x((Fx \vee \phi x) \to (Fx \leftarrow \phi x))$, then the fact that $\Box_F \forall x((Fx \vee \phi x) \to (Fx \leftarrow \phi x))$ grounds itself. This, I take it, is an unwelcome result for Rosen, who asserts that grounding is irreflexive.

For these reasons, I conclude that if real definitions are non-circular, then they are unsubstitutable.

Works Cited

Fine, Kit 1995. Ontological Dependence. *Proceedings of the Aristotelian Society.* Vol 95, 269-290.

Rosen, Gideon 2015. Real Definition. *Analytic Philosophy.* Vol 56, No 3. 189-209