Philos Stud (2022) 179:3019–3036
https://doi.org/10.1007/s11098-022-01812-7

Definition

Samuel Elgin[1]

Accepted: 13 February 2022 / Published online: 11 April 2022

Abstract This paper presents a puzzle about the logic of real definition. I demonstrate that five principles concerning definition—that it is coextensional and irreflexive, that it applies to its cases, that it permits expansion, and that it is itself defined—are logically incompatible. I then explore the advantages and disadvantages of each principle—one of which must be rejected to restore consistency.

Keywords Real definition · Essence · Logic

1 Introduction

Since the inception of our discipline, the notion of real definition has occupied a central role—and in no field is its significance more manifest than metaphysics. Debates within ethics, epistemology, and beyond can all be framed as searches for definitions. For, when the ethicist provides a theory of the good, this can be reasonably understood as a putative definition of the good, and when the epistemologist provides a theory of knowledge, this can be reasonably understood as a putative definition of knowledge. In metaphysics, too, definition plays this role; we might describe a theory of personhood as a view regarding the definition of *being a person*, and a theory of modality as a putative definition of necessity and possibility. But in metaphysics alone definition plays not only this external role—as something that characterizes theories or accounts under consideration—but also an internal role: as an object worthy of investigation itself.

This is not to say that the notion of definition has gone unopposed: far from it. There are any number of reasons why philosophers might object to the framing of our field in

✉ Samuel Elgin
elgin.samuel@gmail.com

[1] University of California San Diego, 9500 Gillman Drive, La Jolla, CA 92093, USA

this way. Perhaps some believe that our theories are too varied for definition to unify them in any theoretically interesting sense; perhaps reality is too coarse-grained for definition to make the kinds of distinctions that metaphysicians typically take it to make; or perhaps reifying definition builds in a gratuitous and suspect piece of ontology. One objection, the response to which constitutes the subject matter of this paper, is that talk of 'definition' is *unintelligible*. I do not hope to assuage all philosophers caught in the grip of the intelligibility concern—as Lewis aptly said, "Any competent philosopher who does not understand something will take care not to understand anything else whereby it might be explained."[1] But it is my aim to provide clarity where there is presently none: to bring structure to the landscape and, in so doing, uncover a puzzle that has thus far remained hidden. For 'definition,' as it is used by the metaphysician, is (among other things) a theoretical term. As such, one way to investigate it systematically is to uncover the theoretical role that it plays.

This strategy involves applying a received view about the introduction of theoretical terms to 'definition.' This view traces back to the Carnap (1958, 1966) discussion of *Ramsey Sentences*, and was given new life by Lewis (1970). The underlying thought is that a theory containing a new predicate is equivalent to its expanded postulate: the claim that there exists a unique *F* that performs every function that the predicate is taken to perform within that theory. In the present case, the disagreement between the adherents and skeptics of a theory of real definition can be understood as a disagreement about whether an *F* exists that theoretically functions as definition is postulated to.

This reframing takes us only so far. Even after a dispute is recognized to be a disagreement over whether a theoretical role is satisfied, it remains unclear how to adjudicate that disagreement. In the sciences, empirical factors often come into play. But in metaphysics, empirical evidence often seems less relevant. Thought experiments bear on the theory of personal identity—laboratory experiments do not. There is, however, an iron weapon within the skeptic's arsenal; if it can be shown that a theory's expanded postulate is logically inconsistent, then the skeptic has won. At that fatal point, there are two ways to respond. One might abandon the theory wholesale and adopt an alternative in its place, or—more modestly—one might embrace a consistent fragment of the original theory.[2] The task for the adherent, on the second strategy, is to determine which consistent fragment to embrace.

It is my claim that this is the status of 'real definition.' Once an expanded postulate is constructed for the theory of definition, it can be shown to be logically inconsistent. The available responses are either to reject that theory entirely or else to embrace a consistent fragment of it. The bulk of this paper concerns the identification of that fragment: the arguments for and against the principles in conflict—one of which must be rejected to restore consistency.

[1] See Lewis (1986).

[2] This, I take it, was the strategy advanced by Dorr and Hawthorne (2013) regarding Lewis's notion of relative naturalness. In practice, of course, there are more options than these two. Perhaps an adherent will claim that there is some ambiguity within the expanded postulate: so that where there appears to be a contradiction, there is in fact none. For the purposes of this paper, I will consider only logically precise theories—ones that do not admit of ambiguity.

The principles at issue could be stated with varying degrees of formalism—ranging from natural language to a logically respectable description. At the outset, I will provide a quasi-formal gloss, as I suspect that it is the most easily intelligible. By the end, I will present a version in terms of a typed, higher-order language—one that is strictly needed to interpret the claim that F and G fall within the extension of 'definition.'[3] Here, '$Def(F, G)$' is to be interpreted as 'F is, by definition, G,' and '$\forall x(Def(Fx, Gx))$' is to be interpreted as 'For all x, the proposition that Fx is, by definition, the proposition that Gx.' The principles are:[4]

Coextensionality:	$Def(F, G) \to \forall x(Fx \leftrightarrow Gx)$
Irreflexivity:	$\neg\exists F(Def(F, F))$
Case congruence:	$Def(F, G) \to \forall x(Def(Fx, Gx))$
Expansion:	$(Def(F, G) \wedge Def(H, I)) \to Def(F, G^{[I/H]})$
Definability:	$\exists F(Def(Def, F))$

The expanded postulate for this theory of definition, then, results from conjoining these five principles and replacing occurrences of '*Def*' with a variable bound by an (higher-order and uniqued) existential quantifier.

Before proceeding to the conflict at hand, it is worth briefly clarifying what these principles mean. *Coextensionality* states that if F is, by definition, G, then F and G are coextensive: an object is F just in case it is G. So, for example, if to be a triangle is, by definition, to be a three-angled polygon, then an object is a triangle just in case it is a three-angled polygon. There are no triangles that are not three-angled polygons—nor are there three-angled polygons that are not triangles. *Irreflexivity* precludes reflexive definitions. It cannot be that to be a person is, by definition, to be a person, or that justice is, by definition, justice. *Case congruence* claims that definitions apply to their cases. If to be a brother is, by definition, to be a male sibling, then for John to be a brother is, by definition, for John to be a male sibling. And if to be a moral agent is, by definition, to be bound by the categorical imperative, then for Sarah to be a moral agent is, by definition, for Sarah to be bound by the categorical imperative. *Expansion* licenses the substitution of some definitions within the contents of others. If hydrogen is, by definition, the element containing a single proton—and if a proton is, by definition, the subatomic particle composed of two up quarks and a down quark, then hydrogen is, by definition, the element containing a single subatomic particle composed of two up quarks and a down quark. And if $\{2\}$ is, by definition, the set containing only the number 2, and

[3] In particular, the language I will employ has two basic types, e and t for entities and sentences respectively, and for any types τ_1 and $\tau_2 \neq e$, $(\tau_1 \to \tau_2)$ is a type. The predicate *Def* as it occurs in '$Def(F, G)$' (for F and G of type $(e \to t)$) is of type $((e \to t) \to ((e \to t) \to t))$. One reason to include the higher-order version of this puzzle is that it removes any concern that this puzzle trades upon a type mismatch between the various principles.

[4] Within this paper, I also assume that classical logic holds (except that I remain neutral on Leibniz's Law). I doubt that that assumption is responsible for this conflict; all inferences used to derive the contradiction are admissible on an intuitionist logic, and while a free logic blocks the penultimate inference, I see no independent reason to adopt a free logic in this context.

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. *Definability*, lastly, states that there exists a definition of real definition—without taking a stand on what the content of that definition is. It asserts that there is some definition of definition or other; definition is not itself a primitive.

I will also assume that polyadic extensions of these principles hold. For example, I will assume that for binary relations R and R', $Def(R, R') \rightarrow \forall x, y(R(x, y) \leftrightarrow R'(x, y))$—an extension of *coextensionality*. If to be a brother is, by definition, to be a male sibling, then Jack is a brother of Jill just in case he is a male sibling of Jill. Relatedly, to extend *case congruence*, I assume that for binary R and R', $Def(R, R') \rightarrow \forall x, y(Def(Rxy, R'xy))$. If to be next to is, by definition, to be adjacent to, then for Jack to be next to Jill is, by definition, for Jack to be adjacent to Jill.

I take it that the commitment to these principles is widespread. As we shall see, this commitment is sometimes made explicit; often, it manifests in practice. Moreover, I have no doubt that many would add further criteria to their preferred expanded postulate: criteria reflecting any additional theoretical work that metaphysicians take definition to perform. But it is enough to begin.

Let 'D' represent the content of the definition of definition—whatever that content might be–and select an arbitrary F and G such that F is, by definition, G.[5] The inconsistency between these principles is brought about in the following way:

(i)	$Def(F, G)$	Supposition
(ii)	$Def(Def, D)$	Definability
(iii)	$Def(Def(F, G), D(F, G))$	*ii*, Case Congruence
(iv)	$Def(Def(F, G), D(G, G))$	*i, iii* Expansion
(v)	$Def(F, G) \leftrightarrow D(G, G)$	*iv*, Coextensionality
(vi)	$D(G, G)$	*i, v* Classical Logic
(vii)	$Def(G, G) \leftrightarrow D(G, G)$	*ii*, Coextensionality
(viii)	$Def(G, G)$	*vi*, *vii*, Classical Logic
(ix)	$\exists H(Def(H, H))$	*viii*, Classical Logic
(x)	$\bot$	*ix*, Irreflexivity

Notably, because the selection of F and G was arbitrary, it not only follows that there is a single, isolated violation of *irreflexivity*. Rather, if *coextensionality, case congruence, expansion,* and *definability* are true, then *every* definition can be used to generate a reflexive definition.[6]

[5] Note that, given *definability*, there must be at least one such case. This puzzle may be generated purely by allowing '$Def(Def, D)$' to witness the schema '$Def(F, G)$' in line *i* in the following derivation.

[6] This falls short of the claim that definition is a reflexive relation. If a term F does not figure as the content of a definition, it may be false that F is, by definition, F. My thanks to Bruno Whittle, who pointed out to me that this argument generalizes beyond an isolated case.

The expanded postulate for this theory of definition is logically inconsistent, and is therefore false. Those who would continue to operate with a notion of definition must identify which part of the theory they reject—i.e., they must identify at least one of the five principles to abandon—and provide a justification for doing so.

Of course, one argumentative technique is apparent; anyone who accepted four of these principles could employ them to derive the negation of the fifth. But that is no help in determining which four to select. What we seek are independent considerations—ones entirely unrelated to this puzzle—that can guide our hand in determining what to do. It is the discussion of these considerations that will occupy the remainder of this paper. For what it's worth, I suspect that many metaphysicians will be loath to reject *coextensionality* and *irreflexivity*; they are starting points in a theory of real definition. However, I ultimately take no stand on how this puzzle ought to be resolved. What I offer are the advantages and disadvantages of each principle. How to weigh these competing considerations is a task that I ultimately leave to the reader.

2 Coextensionality

Coextensionality is the claim that if *F* is, by definition, *G*, then an object is *F* just in case it is *G*. If to believe that *p* is, by definition, to be disposed to act as if *p*, then there are neither cases in which an agent believes that *p* yet is not disposed to act as if *p*, nor cases in which an agent is disposed to act as if *p* yet does not believe that *p*.

There are several reasons to maintain that *coextensionality* is true. Perhaps the most persuasive is an appeal to philosophical practice. Philosophers regularly dismiss putative definitions on the basis of counterexamples. For instance, many deny that knowledge is, by definition, justified true belief, on the grounds that there are plausible cases of justified true belief that are not cases of knowledge.[7] If *coextensionality* were false, it is unclear why this would be the case: why it is a counterexample would have the power to falsify a putative definition. Without *coextensionality*, it could be that knowledge is, by definition, justified true belief even though *knowledge* and *justified true belief* are not coextensive.

Another route to *coextensionality* passes through identity. Many hold that if *F* is, by definition, *G*, then *being F* is the same as *being G* (a view that I will henceforth refer to as 'The Identification Hypothesis').[8] Definition is often held to be reductive; if water is, by definition, the chemical compound H_2O, then water is identical to the chemical compound H_2O. Any account of definition that denies the Identification Hypothesis arguably falls short of these reductive ambitions. But if the Identification Hypothesis is true, then Leibniz's Law comes into play.[9] That is, if *F* is, by definition, *G*, then *F* is the same as *G*, and so they bear all of the same properties. In

[7] See, canonically, Gettier (1963).

[8] See Correia (2017) for someone who assumes that this is true.

[9] Strictly, a higher-order analogue of Leibniz's Law comes into play. There have recently been numerous discussions of higher-order systems that abandon Leibniz's Law: see Caie, Goodman and Lederman (2020), Bacon and Russell (2019), and Bacon (2019).

particular, each bears the property *contains object o within its extension* just in case the other does—and similarly so for all other objects. And, for this reason, *F* and *G* are coextensive.

The Identification Hypothesis provides a path to *coextensionality*—but a controversial one. As we shall see, it restricts the available responses to the puzzle at hand (in particular, it impacts *irreflexivity* and *expansion*). While it does not determine which principle ought to be rejected itself, it has widespread ramifications for any theory of definition. Moreover, it is a hypothesis that some metaphysicians are independently happy to reject.[10] And so, for the moment I simply flag it as a pivotal juncture point.

It is worth pausing to consider how weak a commitment *coextensionality* is. The metaphysical orthodoxy is that definition is co-*in*tensional. That is, if *F* is, by definition, *G*, then *F* and *G* have the same extension in every possible world.[11] Cointensionality is strictly stronger than *coextensionality* (at least if we assume the *T* axiom: $\Box P \rightarrow P$), so those who subscribe to the received wisdom must maintain that *coextensionality* is true. I also note that the general form of *coextensionality* follows from its propositional instance (according to which if *p* is, by definition, *q*, then *p* holds iff *q* holds) and *case congruence*. To see why this is so, take an arbitrary *F* and *G* such that *F* is, by definition, *G*—and an arbitrary object *a*. *Case congruence* entails that *Fa* is, by definition, *Ga*, and the propositional instance of *coextensionality* then entails $Fa \leftrightarrow Ga$. Because the selection of *a* was arbitrary, *F* and *G* are coextensive. Those committed to *case congruence* and the propositional instance of *coextensionality* are thus committed to *coextensionality* in its full generality.

If there are independent reasons to reject *coextensionality*, I am not aware of them.

3 Irreflexivity

Irreflexivity is the claim that there are no reflexive definitions. It is not the case that justice is, by definition, justice or that wisdom is, by definition, wisdom.

Some endorse *irreflexivity* because they maintain that real definition is itself defined in terms of another irreflexive relation. In various ways, Rosen (2015), Correia (2017), and Horvath (2017) each propose a definition of definition in terms of grounding: an asymmetric relation of metaphysical dependence.[12] A bit roughly,

[10] See, e.g., Rosen (2015).

[11] It was once widely held that this conditional could be strengthened into a biconditional: that is, *F* is, by definition, *G*, iff *F* and *G* have the same extension in every possible world. Following examples provided by Fine (1994, 1995a), many maintain that there are necessary connections between properties that are not definitions. Nevertheless, these examples do not undermine the conditional above.

[12] Correia also proposes an account in terms of relative naturalness. What follows is a rough gloss on their views that warrant further refinement. For example, while Rosen takes grounding to be a relation between facts, Correia holds that it is a relation between generics. I direct those interested in the details of these accounts to the original papers.

if *F* is, by definition, *G*, then the fact that *Fa* is grounded in the fact that *Ga*. For example, if to be morally right is, by definition, to maximize utility, then the fact that an act maximizes utility grounds the fact that it is morally right. Because no facts ground themselves, definition is an irreflexive relation; no property is defined in terms of itself. Notably, each author takes the irreflexivity of definition to be not only a feature but a virtue. That is to say, the fact that the resulting accounts preclude reflexive definitions is interpreted as a mark in favor of those very accounts. This strongly indicates that the commitment to *irreflexivity* runs deep.

Additionally, *irreflexivity* 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 is—and if hydrogen is, by definition, the element with a single proton, then protons are more fundamental than hydrogen is. 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* is.[13] Given the (not unreasonable) assumption that nothing is more fundamental than itself, it follows that there are no reflexive definitions.

I suspect that most contemporary philosophers who reject *irreflexivity* will do so because they maintain that identity performs the theoretical work often attributed to definition. This might be understood as the Identification Hypothesis on steroids; not only do definitions entail their corresponding identifications, but rather definitions *simply are* identifications. On this view, the claim '*F* is, by definition, *G*' amounts to the claim 'To be *F* is to be *G*.' Of course, there may be pragmatic reasons to refrain from uttering claims of the form '*F* is, by definition, *F*.' Just as it is infelicitous to respond to 'Who is James?' with 'James is James,' so too it is infelicitous to respond to 'What is virtue?' with 'Virtue is virtue.' But in both cases the answers, though entirely unhelpful, remain strictly true. And so, rather than maintaining that there are *no* reflexive definitions, it might be argued that *everything* can be defined reflexively.

It is not entirely clear how to make this objection stick. The defender of definition is free to grant that there is a reflexive and symmetric reading of 'To be *F* is to be *G*,' but insist that that is not the same reading as intended by their use of '*F* is, by definition, *G*.' Such a metaphysician may claim that their use of 'definition' refers to the subset of identity claims that are substantive—and it is a requirement on substantiveness that the sentence not be reflexive.[14] On this use of 'definition,' the commitment to *irreflexivity* isn't a pragmatic matter at all, but rather a semantic one. And it is far from clear what prevents the metaphysician from using the term 'definition' in that way.

The Identification Hypothesis provides further reasons to reject *irreflexivity*. Suppose that it is correct—i.e., suppose that '*F* is, by definition, *G*' entails 'To be *F* is to be *G*.' If Leibniz's Law holds, this entails that every occurrence of *G* may be

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

[14] For a critique of the literature on identification along these lines, see Cameron (2014).

replaced by an occurrence of *F*; because *F* is the same as *G*, each may be substituted for the other *salva veritate*. But one such substitution results in '*F* is, by definition, *F*'—a reflexive definition. And so, those who accept both the Identification Hypothesis and Leibniz's Law must reject *irreflexivity*. Indeed, it is precisely for this reason that those who accept the Identification Hypothesis often maintain that 'definition' is linguistically opaque and, for this reason, that Leibniz's Law does not apply.[15] Nevertheless, I suspect that sufficiently many will be tempted neither by the rejection of *coextensionality* nor *irreflexivity* that it is worth directing our attention to *case congruence*.

4 Case congruence

Case congruence is the claim that definitions apply to their cases. If to be even is, by definition, to be an integer divisible by two without remainder, then for four to be even is, by definition, for four to be an integer divisible by two without remainder, and if to be a béchamel is, by definition, to be a roux with milk, then for sauce *s* to be a béchamel is, by definition, for sauce *s* to be a roux with milk.

As with *coextensionality*, an initial defense of *case congruence* is made by appeal to practice. If it were false, then it ought to admit of counterexample. It may be, for instance, that to be morally right is, by definition, to maximize utility, and for Tim's act to be morally right is, by definition, for Tim's act to comply with the categorical imperative. I am aware of no philosophers who have made claims along these lines—and I take this to indicate that the tacit commitment to *case congruence* is widespread.[16]

Those who would resolve the present dilemma by rejecting *case congruence* might accept an alternative in its place.[17] Roughly, the thought is that only logical simples have definitions. While it may be that to be even is, by definition, to be divisible by two without remainder, the proposition that four is even lacks a definition—it itself is primitive. For this reason, claims like 'To be morally right is, by definition, to maximize utility, and for Tim's act to be morally right is, by definition, for Tim's act to comply with the categorical imperative' are universally

[15] For example, see Correia (2017).

[16] This particular example is somewhat tricky. Parfit (2011) argues that Kantianism and consequentialism (as well as contractarianism) are unified in the sense that the best versions of the three views are the same—advocates for each have been climbing the same mountain from different sides. A Parfitian might claim something close to the example above: the definition of the right is given in consequentialist terms and the definition of Tim's act being morally right is given in Kantian terms, because consequentialism and Kantianism are one and the same (at least when understood correctly). But this is not a counterexample to *case congruence*—precisely because the Parfitian identifies the consequentialist view with the Kantian view. A counterexample would be someone who maintains that the two are distinct, rival views—and while the right is defined in consequentialist terms, the claim that Tim's act is right is defined in Kantian terms. I know of no one who subscribes to such a claim.

[17] My thanks to Jeremy Goodman for this suggestion.

false. Because ‘for Tim’s act to be morally right’ is logically complex, it is not the sort of thing that has a definition. And so, while this proposal denies *case congruence*, it doesn’t allow for different definitions of the right to figure in generic and specific cases. While the right may itself be defined, any particular instance concerning a right act is logically complex, and so inapt for definition.

Along these lines, we may then define a relation of *definitional equivalence*. If a logically complex term contains a simple with a definition, then that term is definitionally equivalent to the result of replacing that constituent with its definition. For example, if ‘To be a bachelor is, by definition, to be an unmarried male,’ then ‘For John to be a bachelor’ is definitionally equivalent to ‘For John to be an unmarried male.’ Definitional equivalence, unlike definition, is a reflexive relation: everything is definitionally equivalent to itself. On this view, the conflict regarding definition is to be resolved by rejecting *case congruence*, and the analogous conflict for definitional equivalence is to be resolved by rejecting *irreflexivity*.

5 Expansion

Expansion is the claim that, within the content of a definition, terms may be replaced by their own definitions. For example, if {Socrates} is, by definition, the set containing only Socrates, and Socrates is, by definition, the result of *this* sperm and *that* egg, then {Socrates} is, by definition, the set containing only the result of *this* sperm and *that* egg. And if to be even is, by definition, to be a natural number divisible by two without remainder, and two is, by definition, the successor to the number one, then to be even is, by definition, to be a natural number divisible by the successor to the number one without remainder.

Expansion is a restricted substitution principle. It permits substitution within the definiens—or content of definition—but not the definiendum—or object being defined. This restriction matters because an unrestricted principle (i.e., one that allowed for substitution within both the content and object of definition) immediately generates reflexive definitions. Consider the following unrestricted principle:

A is, by definition, *B*
C is, by definition, *D*

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

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 permissible. 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. Quite generally, by allowing the same example to witness the first two conditions, it is possible to derive reflexive definitions.[18] *Expansion*, as it is stated, does not license substitution within the definiendum, so it does not conflict with *irreflexivity* in this way.

Once again, the Identification Hypothesis rears its head. If it holds, and if 'definition' is transparent, then *expansion* holds as well. Given the Identification Hypothesis, if *F* is, by definition, *G*, then *F* is the same as *G*. Leibniz's Law then permits the substitution of *F* for *G* in any context, including the contents of other definitions. However, as previously noted, the conjunction of the Identification Hypothesis and Leibniz's Law already forces the rejection of *irreflexivity*, so anyone who adopts both principles has no additional reason to reject *expansion*—or, indeed, any other principle under discussion.

Expansion is the cousin of *transitivity*—the claim that if *A* is, by definition, *B,* and *B* is, by definition, *C*, then *A* is, by definition, *C*. Strictly, *expansion* is stronger than *transitivity*; *transitivity* can be considered as the limiting case of *expansion* in which the term being substituted for is the definiens in its entirety. But while *expansion* allows us to 'dive into' the content of the definiens and replace some terms with others, *transitivity* does not—it applies only to the definiens in its entirety. As such, while *expansion* entails *transitivity*, *transitivity* does not entail *expansion*.

The commitment to *transitivity* is widespread—though typically given without argument.[19] It is often taken to be a starting point in a theory of definition; it is considered a mark in favor of a theory if it can be shown to be transitive. One path to *expansion* appeals to *transitivity*. For, while one can consistently hold that *transitivity* is true while *expansion* is false, it is not at all clear why we should do so: why we should expect *transitivity* to succeed and *expansion* to fail.[20]

The explicit commitment to *expansion* is less common than the commitment to *transitivity* (though its explicit denial is, as far as I know, nonexistent). An exception to this general rule is the following:

> "It should be possible to prove a principle that licenses arbitrary definitional expansion:
>
> $Def(F, \Phi)$ and $Def(G, \Psi)$ then $Def(F, \Phi^{\Psi/G})$

[18] Note that this is slightly weaker than the claim that definition is *reflexive*—i.e., from the claim that every instance of '*A* is, by definition, *A*' is true. There may well be some irreflexive cases; what this principle entails is that reflexivity arises for every term that serves as the content of a definition. There is a model on which both this substitution principle and *irreflexivity* are true—a model in which everything is primitive, and nothing has a definition.

[19] See, most explicitly, Correia (2017), Rosen (2015), and Horvath (2017). For endorsements of the transitivity of related phenomena such as ground, essence, and relative fundamentality, see, for example, Fine (1995b, 2012), deRosset (2013, 2017), Dasgupta (2016), Berker (2017), and Dixon (2018). The closest thing to an explicit disavowal of *transitivity* occurs in Schaffer (2012). For a reply, see Litland (2013).

[20] In addition, much theoretical work attributed to *transitivity* can only be adequately accomplished by *expansion*. For example, one type of ontological dependence can be understood in terms of definitional containment—see Fine (1995a). Entity *e* ontologically depends upon entity *e*′ just in case *e*′ figures within the definition of *e*. *Expansion* can be used to derive the transitivity of ontological dependence, but *transitivity* is strictly compatible with the claim that ontological dependence is intransitive.

> Where $\Phi^{\Psi/G}$ is the result of substituting Ψ for G in Φ...Any account of real definition should license the substitution of definiens for definiendum in a ground to yield a further ground" (Rosen 2015, pg. 201).[21]

Notably, Rosen claims not only that *expansion* holds but that it ought to be *provable* that it holds. This suggests a path toward *expansion*; we ought to believe it because of its proof. To the best of my knowledge, however, *expansion* does not follow from any of the widely accepted principles about essence or definition. Minimally, I have been unable to prove it from them. Those seeking a proof of *expansion* should look—not to the logic of essence and definition—but to the logic of identity.

Within the recent (and rapidly expanding) literature on higher-order identity, Caie, Goodman, and Lederman (2020) provide a proof of Leibniz's Law. The aim of this proof is not to vindicate Leibniz's Law, but rather to systematically investigate which principles must be abandoned in languages with opaque predicates. As it turns out, this derivation can be modified to prove *expansion*.

Let us adopt a typed, higher-order language with λ-abstraction. Within this language, there are two basic types e, t for the type of entities and sentences respectively, and for any types τ_1, and $\tau_2 \neq e$, $(\tau_1 \rightarrow \tau_2)$ is a type; nothing else is a type. Monadic first-order predicates can be identified as terms of type $(e \rightarrow t)$, diadic first-order predicates are terms of type $(e \rightarrow (e \rightarrow t))$, etc. Monadic second-order predicates are of type $((e \rightarrow t) \rightarrow t)$, and monadic third-order predicates are of type $(((e \rightarrow t) \rightarrow t) \rightarrow t)$. The negation operator $\neg$ is of type $(t \rightarrow t)$, and the binary operators $\wedge, \vee, \rightarrow, \leftrightarrow$ are all of type $(t \rightarrow (t \rightarrow t))$. Additionally, this language is equipped with infinitely many variables of every type, as well as the quantifiers $\exists, \forall$ of types $((\tau \rightarrow t) \rightarrow t)$ (for every type τ).

In first-order languages, these quantifiers perform dual functions. They serve both to express generality and to bind the variables occurring within their scope. But in higher-order languages, these tasks are divided: the task of expressing generality is performed by quantifiers, and the task of variable binding is performed solely by the λ-terms. Thus, 'There exists an F' is expressed as '$\exists\lambda x.(Fx)$,' rather than '$\exists x(Fx)$.'

Lastly, for each type τ there exists a predicate *Def* of type $(\tau \rightarrow (\tau \rightarrow t))$ that is used to express definitions. The intended interpretation of $\ulcorner Def^{(\tau\rightarrow(\tau\rightarrow t))}(A^\tau, B^\tau)\urcorner$ is 'A is, by definition, B.'

The principles that generate *expansion* (which are to be read either as schemata with applications in every type, or else as terms whose type is contextually evident) are the following:

Material abstraction	$Def(\phi, \psi) \rightarrow Def(\lambda x.\phi[x/a], \lambda x.\psi[x/a])$
Application congruence	$Def(F, G) \wedge Def(a, b) \rightarrow Def(Fa, Gb)$
Beta-Eta equivalence	ϕ may be replaced by ψ provided ϕ and ψ are $\beta\eta$ equivalent

[21] For another commitment to this type of principle, see Correia and Skiles (2019).

The derivation of *expansion* proceeds as follows:

(i)	$Def(a, b)$	Supposition
(ii)	$Def(\phi, \psi)$	Supposition
(iii)	$Def(\lambda x.\phi[x/a], \lambda x.\psi[x/a])$	ii, Material abstraction
(iv)	$Def(\lambda x.\phi[x/a](a), \lambda x.\psi[x/a](b))$	iii, Application congruence
(v)	$Def(\phi, \psi[b/a])$	iv, βη equivalence

Therefore, if *material abstraction, application congruence,* and *βη equivalence* are all true, then *expansion* is true as well. Those who would reject *expansion* must also reject (at least) one of these three principles.

It's worth clarifying what $\beta\eta$ equivalence amounts to, because there are several related—yet importantly distinct—principles in this area. One is a principle of conversion. It allows one to infer Fa from $\lambda x.Fx(a)$—one may infer the proposition that Fa from the lambda-abstract for *being* F as applied to a. To the best of my knowledge, this is uncontroversial (indeed, it might be considered constitutive of the higher-order languages used in this derivation).[22] Another—and much more controversial—principle is that identity is preserved through this conversion. That is to say, not only may one infer that Fa from $\lambda x.Fx(a)$, but the two sentences express the very same proposition.[23]

The principle of $\beta\eta$ equivalence at issue is neither of these. It is stronger than the inferential claim yet weaker than the identity. What is required is that if there are two expressions that are $\beta\eta$ equivalent, then one may be substituted for the other. The ability to infer one expression from the other does itself not guarantee this substitution. In contrast, the identity principle *does* entail this substitution principle—at least when combined with Leibniz's Law. After all, Leibniz's Law states that terms that denote the same thing may be substituted for one another, and the principle of identity states that $\beta\eta$-equivalent expressions denote the same thing. But the substitution principle need not presuppose identity. What is required only is that one expression may take the other's place; we need take no stand on whether they denote the same thing.

Application congruence allows for the combination of two definitions into one. If to be human is, by definition, to be a rational animal, and Aristotle is, by definition, the result of *this* sperm and *that* egg, then for Aristotle to be human is, by definition,

---

22 I note, however, that there may be some who object even to this very weak principle in contexts with opaque predicates.

23 For a discussion of this point, see Dorr (2016). It is controversial because it rules out the view that propositions are structured. In particular, when combined with the claim that $Fa = Gb \rightarrow F = G$ (which is integral to the structured-proposition account), it has the untenable result that propositions with different extensions are identical. To see why this is so, note that $\beta\eta$ conversion entails $\lambda x.Rxx(a) = \lambda x.Rxa(a)$: the proposition that object a stands in relation R to itself is the same as the proposition that object a stands in relation R to a. On the structured proposition view, this entails that $\lambda x.Rxx = \lambda x.Rxa$: the property of standing in relation R to oneself is identical to the property of standing in relation R to a. These properties need not be coextensive: one may be as tall as oneself without being as tall as Shaquille O'Neal. While Dorr takes this to count against structured propositions, others may reject the claim that identity is preserved through $\beta\eta$ conversion.

for the result of *this* sperm and *that* egg to be a rational animal. *Application congruence* strongly resembles *case congruence*, and many reasons to accept (and reject) *case congruence* apply to *application congruence* as well. For the moment, suffice it to say that I can think of no plausible instances in which it fails.

Material abstraction is the near converse of *case congruence*. Just as *case congruence* allows one to infer that 'For Linda to be a sister is, by definition, for Linda to be a female sibling' from 'To be a sister is, by definition, to be a female sibling,' so too *material abstraction* allows one to infer 'To be a sister is, by definition, to be a female sibling' from 'For Linda to be a sister is, by definition, for Linda to be a female sibling.' The underlying thought is that when a term appears in both the definiendum and definiens—within both the object and content of analysis—then that term is not responsible for the definition in question. That is to say, there is a plausible *non-circularity* criterion on definition. While terms can (and do) appear in both the object and contents of definitions, they cannot appear *essentially* in both the object and content—they are not the reason a given expression constitutes a definition.[24] And, because these terms are inessential, they can be abstracted away.

As stated, however, *material abstraction* appears problematic. Suppose there were a p and ϕ such that p occurs nowhere in ϕ and $Def(p, \phi)$. In this case, *material abstraction* would entail that $Def(\lambda x.x, \lambda x.\phi)$—as *material abstraction* allows us to abstract away the entirety of the object being defined, and yet there is nothing to abstract away from the content of definition. One response would be to restrict *material abstraction* so that the terms abstracted occur both in the object and content of definition (that is, the principle may be restricted to preclude vacuous abstraction). But, even with this restriction in place, *material abstraction* remains controversial.[25] Perhaps there are some cases where matter occurs essentially within a definition, and in other cases it occurs accidentally. *material abstraction* makes no such distinction and allows the abstraction of any material: accidental or essential.

To clarify the worry, suppose there is a property F and relation R such that $Def(F, \lambda x.Rxa)$. *Being F* is, by definition, standing in relation R to object a. Using various principles referenced throughout this paper, we may derive that F is, by definition, standing in relation R to oneself in the following way:

(i)	$Def(F, \lambda x.Rxa)$	Supposition
(ii)	$Def(Fa, \lambda x.Rxx(a))$	i, Case congruence
(iii)	$Def(Fa, Raa)$	ii, $\beta\eta$ Equivalence
(iv)	$Def(\lambda x.Fx, \lambda x.Rxx)$	iii, Material abstraction
(v)	$Def(F, \lambda x.Rxx)$	iv, $\beta\eta$ Equivalence

[24] At least, they can appear in both the object and content of definition on the assumption that *case congruence* is true.

[25] My thanks to an anonymous reviewer for this suggestion.

And some might reasonably maintain that although F is, by definition, standing in relation R to a, it is not, by definition, standing in relation R to oneself. For this reason, there is an independent motivation to reject *material abstraction.*

To my mind, the defenses of both *application congruence* and *material abstraction* are defeasible. They are not knock-down considerations. One path to the resolution of the problem at issue is the rejection of *expansion*. However, this rejection must be accompanied by the rejection of *application congruence*, *material abstraction,* or *beta-eta equivalence*, as these principles entail that *expansion* is true.

6 Definability

Definability is the claim that there exists a definition of definition. Definition does not rank among the primitive relations—it is defined in terms of something or other. I suspect that (at least to some) this principle seems relatively controversial. On one interpretation, definition forms a bedrock of our discipline: a foundation upon which other philosophical accounts rest. And so, the contention that definition is itself primitive is not entirely implausible. Moreover, while many of the previous principles appeared to be implicit in philosophical practice, this is not so for *definability*. There is no reason to suspect that ethicists, epistemologists, and the like tacitly assume that real definition is itself defined.

Nevertheless, numerous philosophers maintain that *definability* is true. Typically, this occurs because philosophers provide an account of definition.[26] Philosophers defend a particular view about what the definition of definition is, and are thereby committed to the claim that definition has some definition or other. Correspondingly, one defense of *definability* is parasitic on any argument that they provide. Any reason to support their views constitutes a reason to endorse *definability*.

I believe that there is a further reason to support *definability*—but one that is (or at least ought to be) controversial.

Definitions provide answers to metaphysical-why questions. Several things might be intended by a question like 'Why is Fred a bachelor?' Often, it might be used to enquire into the reason for Fred's marital status. In these cases, responses like 'Because he has not yet fallen in love' are appropriate. But there is a metaphysical reading of this question as well—one concerning what it is in virtue of that Fred counts as a bachelor—and it is here that an appeal to definition can naturally be made. To the metaphysician, the response 'Because Fred is an unmarried male, and to be a bachelor is, by definition, to be an unmarried male' seems as satisfying an answer as any.

In a similar manner, the definition of definition provides an answer to metaphysical-why questions. Let us suppose, for the sake of a concrete example, that to be morally right is, by definition, to maximize utility. It is reasonable to enquire *why* the right is defined as it is: what makes it the case that the right is

[26] As before, see Rosen (2015), Correia (2017), and Horvath (2017). I direct those interested in the content of these views to their original papers.

defined in terms of that which maximizes utility rather than that which cultivates the virtues. And just as the answer to 'Why is Fred a bachelor?' naturally appeals to the definition of *being a bachelor*, so too the answer to, 'Why is the right, by definition, maximizing utility?' naturally appeals to the definition of definition. The reason the right is defined in terms of maximizing utility is that it stands in the appropriate relation to maximizing utility: a relation given by the definition of definition.

Those who accept *definability* have the resources to metaphysically explain why it is properties and relations are defined as they are; they can appeal to the definition of definition to provide such an account. In contrast, those who reject *definability* cannot respond in this way.[27] And so, one reason to accept *definability*—beyond the appeal of particular accounts of definition—is that it provides resources for metaphysical explanations that we seek.

There is, however, a reason to reject *definability*, one so initially compelling that it suggests that the preceding discussion ought to have been curtailed. As stated, *definability* claims that a relation—in particular, the relation of definition—falls within its own extension (while remaining agnostic as to what it stands in that relation to). But there is a strong reason to deny that any property or relation falls within its own extension: the Russell Paradox. For, if properties fall within their own extension, it is natural to maintain that there is a property of *being a property that does not fall within its own extension*: a property that falls within its own extension just in case it does not. This problem can be avoided by denying that properties are the types of things that can fall within their own extension. And if *no* property or relation falls within its own extension, then definition does not fall within the extension of definition, and so we ought to reject *definability*.

This, as I said, is an extraordinarily compelling point. It is also false. Of course, there are numerous ways we might attempt to avoid the Russell Paradox, but the *obvious* method leaves the present puzzle intact. While outright contradiction is avoided, the conflict between the five principles at issue remains. What is this obvious method? To adopt a typed higher-order language in which the claim that a property falls within its own extension is strictly ungrammatical, and so inapt for truth or falsity.

Fortunately, we have already encountered such a language, so no new formalism is required. As before, let us assume that there are two basic types, e and t (for the types of entities and sentences respectively), and that for any types $\tau_1, \tau_2 \neq e$, $(\tau_1 \rightarrow \tau_2)$ is a type, and nothing else is a type. We allow for infinitely many variables of every type, and the corresponding λ-abstracts needed to bind them—and the quantifiers $\forall, \exists$. Furthermore, for any type τ, there exists a predicate *Def* of type $\tau \rightarrow (\tau \rightarrow t)$ with the intended interpretation that $\ulcorner Def^{\tau\rightarrow(\tau\rightarrow t)}(A^{\tau}, B^{\tau})\urcorner$ asserts that A is, by definition, B. Because this language is typed, the Russell Paradox is avoided. The only additional symbolism—which I introduce solely to reduce the length of types in the principles and subsequent derivation—is τ^2 (for a generic type

[27] Of course, they might be able to offer a different explanation for why properties and relations are defined as they are. But it is incumbent on those who reject *definability* and would seek to metaphysically explain why properties and relations are so defined to indicate what those explanations look like. My thanks to Alexander Skiles for pressing me on this point.

τ), which is shorthand for $\tau \rightarrow (\tau \rightarrow t)$. Relatedly, $(\tau \rightarrow t)^2$ is shorthand for $(\tau \rightarrow t) \rightarrow ((\tau \rightarrow t) \rightarrow t))$.

With this language in place, the five principles at issue can be stated in a logically precise manner. Strictly, these principles become schemata with applications for each type τ. In cases where the type is not explicitly mentioned, it is contextually evident.

Coextensionality:	$Def^{t^2}(P^t, Q^t) \rightarrow (P^t \leftrightarrow Q^t)$
Irreflexivity:	$\neg\exists\lambda X^\tau.Def^{\tau^2}(X, X)$
Case congruence:	$Def^{(\tau\rightarrow t)^2}(F^{\tau\rightarrow t}, G^{\tau\rightarrow t}) \rightarrow Def^{t^2}(F^{\tau\rightarrow t}(a^\tau), G^{\tau\rightarrow t}(a^\tau))$
Expansion:	$(Def^{\tau^2}(F^\tau, G^\tau) \wedge Def^{\alpha^2}(H^\alpha, I^\alpha)) \rightarrow Def^{\tau^2}(F^\tau, G^{\tau[\mathrm{I/H}]})$
Definability:	$\exists\lambda X^{\tau^2}.Def^{\tau^2\rightarrow(\tau^2\rightarrow t)}(Def^{\tau^2}, X)$

Most of these amount to the reframing of the original principles in a paradox-free language. *Definability*, for example, amounts to the claim that there is a higher-order definition for each lower-order definition. *Coextensionality*, however, has been restricted to its propositional instance, rather than for all types generally.[28]

The framing of this puzzle within a typed language offers another potential resource. It may be that different principles are rejected for different types. Perhaps, for example, *definability* is to be rejected for the predicate $Def^{e\rightarrow(e\rightarrow t)}$ while *expansion* is to be rejected for the predicate $Def^{(e\rightarrow t)\rightarrow((e\rightarrow t)\rightarrow t)}$. However, I can think of no reason to reject different principles for different types, so I merely note that it is an option in logical space.

Within this language, the conflict can then be derived as follows:

(i)	$Def^{\tau^2}(F^\tau, G^\tau)$	Supposition
(ii)	$Def^{\tau^2\rightarrow(\tau^2\rightarrow t)}(Def^{\tau^2}, D^{\tau^2})$	Definability
(iii)	$Def^{t^2}(Def^{\tau^2}(F^\tau, G^\tau), D^{\tau^2}(F^\tau, G^\tau))$	*ii*, Case congruence
(iv)	$Def^{t^2}(Def^{\tau^2}(F^\tau, G^\tau), D^{\tau^2}(G^\tau, G^\tau))$	*i*, *iii*, Expansion
(v)	$Def^{\tau^2}(F^\tau, G^\tau) \leftrightarrow D^{\tau^2}(G^\tau, G^\tau)$	*iv*, Coextensionality
(vi)	$D^{\tau^2}(G^\tau, G^\tau)$	*i*, *v*, Classical logic
(vii)	$Def^{t^2}(Def^{\tau^2}(G^\tau, G^\tau), D^{\tau^2}(G^\tau, G^\tau))$	*ii*, Case congruence
(viii)	$Def^{\tau^2}(G^\tau, G^\tau) \leftrightarrow D^{\tau^2}(G^\tau, G^\tau)$	*vii*, Coextensionality
(ix)	$Def^{\tau^2}(G^\tau, G^\tau)$	*vi*, *viii*, Classical logic
(x)	$\exists\lambda X^\tau.Def^{\tau^2}(X, X)$	*ix*, Classical logic
(xi)	$\perp$	*x*, Irreflexivity

[28] Recall that the propositional instance and *case congruence* collectively entail *coextensionality* in its full generality.

The upshot, then, is this: there was a presumptive concern regarding *definability*. It appeared to assert that a relation—in particular the relation of real definition—fell within its own extension. This naturally gives rise to paradox and provides an initial reason to reject *definability*. However, once we shift into a typed language, the threat of paradox is removed, and yet the present conflict remains. And so, if there is a reason to reject *definability*, it is not due to the threat of paradox.

7 Conclusion

I close by returning to where we began: a discussion of the expanded postulate for a theory of definition. I have no doubt that some readers suspect that this expanded postulate (whatever it may be) constitutes the definition of real definition. What definition *is* is that relation that performs the theoretical work attributed to real definition. And so, once we identify what that work consists of, we will thereby have identified what the definition of definition is.

Lewis's (1970) original work suggests that this is incorrect. Note that his account concerns how to define theoretical *terms*, rather than properties and relations. That suggests that it provides nominal definitions, rather than real definitions. The expanded postulate constitutes a nominal definition of 'definition' rather than a real definition of definition. It merely specifies what the word means as used by the metaphysician.

My own view is that matters are not so straightforward. What an expanded postulate is is the formal description of the theoretical function that a property (or relation) performs. Properties or relations that are defined in terms of their expanded postulates are thus those that are functionally defined. If definition is one such property—if it is functionally defined—then its expanded postulate provides its real definition. In contrast, if definition is not functionally defined, then its expanded postulate provides a merely nominal definition. The debate over the relation between definition and its expanded postulate can thus be understood as a debate over whether definition is itself functionally defined.

There is a conflict between the principles *coextensionality, irreflexivity, case congruence, expansion,* and *definability*. Each is initially plausible, and while there are modest reasons to reject some, others garner a great measure of support. Perhaps some will respond by abandoning the theory of definition wholesale. I myself doubt that this is appropriate—a difficulty in rejecting one principle is no reason to reject five. But although I ultimately take no stand on how this puzzle ought to be resolved, something must be done. No contradiction may be allowed to remain.

Acknowledgements My thanks to Michael Della Rocca, Catherine Elgin, Jeremy Goodman, Nathaniel Imel, Will Ratoff, and Ralph Wedgewood for providing comments on earlier drafts of this paper - as well as to the attendees of the 2021 Central Division of the American Philosophical Association's session on Essence, Definition and Philosophical Analysis, and UC San Diego's metaphysics working group for their feedback. I am especially grateful to Cian Dorr, whose correspondence led directly to the development of this paper.

References

Bacon, A. (2019). Substitution structures. *Journal of Philosophical Logic, 48*, 1017–75.
Bacon, A., & Russell, J. (2019). The logic of opacity. *Philosophy and Phenomenological Research, 99*(1), 81–114.
Berker, S. (2017). The unity of grounding. *Mind, 127*(507), 729–77.
Caie, M., Goodman, J., & Lederman, H. (2020). Classical opacity. *Philosophy and Phenomenological Research, 101*(3), 524–66.
Cameron, R. (2014). On the lack of direction in Rayo's the construction of logical space. *Inquiry, 57*(4), 427–41.
Carnap, R. (1958). Observational and theoretical language. *Dialectica, 12*(3–4), 236–48.
Carnap, R. (1966). *Philosophical foundations of physics: An introduction to the philosophy of science*. Basic Books.
Correia, F. (2017). Real definitions. *Philosophical Issues, 27*(1), 52–73.
Correia, F., & Skiles, A. (2019). Grounding, essence and identity. *Philosophy and Phenomenological Research, 3,* 642–70.
Dasgupta, S. (2016). Metaphysical rationalism. *Noûs, 50*(2), 379–418.
deRosset, L. (2013). Grounding explanations. *Philosopher's Imprint, 13*(7), 1–26.
deRosset, L. (2017). Grounding the unreal. *Philosophy and Phenomenological Research, 95*(3), 535–63.
Dixon, S. (2018). Upward grounding. *Philosophy and Phenomenological Research, 97*(1), 48–78.
Dorr, C. (2016). To be F is to be G. *Philosophical Perspectives, 30*(1), 39–134.
Dorr, C., & Hawthorne, J. (2013). Naturalness. In K. Bennett (Ed.) *Oxford Studies in Metaphysics*.
Fine, K. (1994). Essence and modality. *Philosophical Perspectives, 8,* 1–16.
Fine, K. (1995a). Ontological dependence. *Proceedings of the Aristotelian Society, 95,* 269–90.
Fine, K. (1995b). Ontological dependence. *Proceedings of the Aristotelian Society, 95*(1), 269–90.
Fine, K. (2012). A guide to ground. In F. Correia & B. Schnieder (Eds.), *Metaphysical grounding* (pp. 37–80). Cambridge University Press.
Gettier, E. (1963). Is knowledge justified true belief? *Analysis, 6,* 121–3.
Horvath, J. (2017). Philosophical analysis: The concept grounding view. *Philosophy and Phenomenological Research, 97*(3), 724–50.
Lewis, D. (1970). How to define theoretical terms. *Journal of Philosophy, 67*(13), 427–46.
Lewis, D. (1986). *On the plurality of worlds*. Oxford University Press.
Litland, J. (2013). On some counterexamples to the transitivity of ground. *Essays in Philosophy, 14*(1), 19–32.
Parfit, D. (2011). *On what matters*. Oxford University Press.
Rosen, G. (2015). Real definition. *Analytic Philosophy, 56*(3), 189–205.
Schaffer, J. (2012). Grounding, transitivity and contrastivity. In F. Correia & B. Schnieder (Eds.), *Metaphysical grounding: Understanding the structure of reality* (pp. 122–138). Cambridge University Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.