Australasian Journal of Logic

Non-classical circular definitions

Shawn Standefer
The University of Melbourne

Abstract

Circular definitions have primarily been studied in revision theory in the classical scheme. I present systems of circular definitions in the Strong Kleene and supervaluation schemes and provide complete proof systems for them. One class of definitions, the intrinsic definitions, naturally arises in both schemes. I survey some of the features of this class of definitions.

In this paper, I will study circular definitions from two non-classical perspectives, complementing the most developed extant approach, revision theory, which has primarily been executed in the classical scheme. "Circular definition" is often taken narrowly to mean that the *definiendum* appears in the *definiens*. I will use "circular definition," and sometimes just "definition," more broadly here to encompass definitions that are circular in the sense just mentioned, that are not circular but are interdependent, and even ones that fall into neither of those categories.[1] Given a countable first-order language $\mathcal{L}$, the *base language*, expand the language to $\mathcal{L}^+$ by adding at most countably many new predicate letters G_i, each of which receives a definitional clause in the set of definitions $\mathscr{D}$.[2] The set $\mathscr{D}$ may be any set of definitional clauses, as shown in table 1, where A_{G_i} is any formula of $\mathcal{L}^+$, and the only restriction on the definitional clauses is that no variables other than $\bar{x}_i$ may appear free in A_{G_i}.[3]

[1] This usage is in the same spirit as the usage of "partial function" that encompasses the total functions as well as the properly partial ones.

[2] I will use uppercase letters from the middle of the alphabet, such as G, H, and J, for defined predicates of a set of definitions and script letters, $\mathscr{D}, \mathscr{E}$, etc., for sets of definitions.

[3] I will use bar notation for sequences of terms, e.g. $\bar{x}$ for $x_1, \ldots, x_n$ and $\bar{t}$ for $t_1, \ldots, t_m$, or of objects, e.g. $\bar{d}$ for $d_1, \ldots, d_k$.

$$\begin{array}{lll} G_1(\bar{x}_1) & =_{Df} & A_{G_1}(\bar{x}_1) \\ G_2(\bar{x}_2) & =_{Df} & A_{G_2}(\bar{x}_2) \\ & \vdots & \\ G_k(\bar{x}_k) & =_{Df} & A_{G_k}(\bar{x}_k) \\ & \vdots & \end{array}$$

Table 1: The form of circular definitions

The paper will proceed as follows. I will begin presenting two different motivations for investigating circular definitions: via considerations of truth (§1.1) and via considerations of meaning (§1.2). Following that, I will provide some indication of what is gained from looking at circular definitions, in particular (§1.3). I will then provide some motivation for the approach of this paper, namely investigating *non-classical* circular definitions (§1.4). Then I will look at two non-classical approaches to definitions, based on the Strong Kleene scheme (§2) and the supervaluation scheme (§3). For each I will provide sound and complete proof systems for their respective notions of validity. Next, I will highlight one class of definitions that arises naturally from model-theoretic considerations and explore some of its features (§4). Finally, I will close with some directions for future research (§5).

1 Introduction and motivations

There are two roads to circular definitions that I will cover here: via truth, and via a certain conception of meaning.[4] Let us begin with truth.

[4]There are two other ways into circular definitions besides the ones I will focus on. Yablo [1993] considers definitions from the point of view of introducing new predicates into a language along with rules governing their usage. Circular definitions arise in that context via rules that can invoke themselves and other rules. Motivated by considerations from logic programming, Schroeder-Heister [1993] considers proof-theoretic frameworks in which there are rules of definitional reflection that introduce defined atoms. No restrictions are placed on the rules, so they accommodate circular definitions.

1.1 From truth to circular definitions

Saul Kripke, and independently Robert Martin and Peter Woodruff, came up with fixed-point theories of truth.[5] In doing so, Kripke showed how to add to the language of arithmetic an untyped truth predicate for the expanded language without collapsing into triviality on account of the paradoxes. On this approach, the semantic value of the truth predicate is taken to be three-valued, a fixed-point of an operation on three-valued interpretations. Kripke presented a construction of one of these fixed-points as being built up through iterations of an operation on interpretations. One starts with a model for the language with truth and improves the interpretation of the truth predicate by filling out its extension and anti-extension in stages, based on what sentences receive the values 1 and 0, respectively. The construction is bound to eventually reach a fixed-point, a point after which further iterations of the operation yield nothing new. Some sentences do not receive a classical semantic value in this construction, and the resulting fixed-point has truth-value gaps. Kripke's construction works with a range of semantic schemes, although it requires the use of a non-classical scheme. In the case of the Strong Kleene scheme, the truth predicate is transparent, in the sense that for all sentences A and contexts C, $C(A)$ gets the same semantic value as $C(T(\ulcorner A \urcorner))$.[6]

In response to this work on truth, Anil Gupta and, independently, Hans Herzberger developed revision theories for truth.[7] They wanted to develop an approach to an untyped truth predicate that worked with the classical scheme, unlike Kripke's approach. The results were the first revision theories of truth.

Formally, Gupta and Herzberger used sequences of classical, two-valued interpretations for the truth predicate. These sequences are generated by an operation determined by the Tarski biconditionals, $T(\ulcorner A \urcorner)$ iff A, where "iff" is not understood as the material biconditional. Rather, the biconditional is taken to determine how to revise the interpretation of the truth predicate: if A receives the semantic value 1, or 0, at stage α, then $T(\ulcorner A \urcorner)$ receives the semantic value 1, or 0, respectively, at stage $\alpha + 1$. Unlike with Kripke's

[5] See Kripke [1975] and Martin and Woodruff [1975]. Their work bears similarities to that of Gilmore [1974] and Brady [1971] in set theory.

[6] $\ulcorner . \urcorner$ is an operation that yields names of sentences. 'T' is a truth predicate.

[7] See Gupta [1982] and Herzberger [1982], as well as Belnap [1982]. See Gupta and Belnap [1993] for a comprehensive presentation of revision theory.

construction, these sequences need not reach fixed-points, and in fact often do not. Certain sentences, however, stabilize as 1 or 0.

According to Gupta and Belnap [1993], the Tarski biconditionals together provide a *circular definition* of the truth predicate. The definition is circular since the set of Tarski biconditionals for the language containing the truth predicate itself will contain instances with an ineliminable truth predicate occurring in the *definiens*. Take, for example, $T(\ulcorner\forall x{\sim}(Tx\,\&\,{\sim}Tx)\urcorner)$ iff $\forall x{\sim}(Tx\,\&\,{\sim}Tx)$; to evaluate the sentence on the right, one will have to evaluate an instantiation which is the sentence on the left, which in turn requires evaluating the sentence on the right. One, then, cannot eliminate all occurrences of the truth predicate. Gupta and Belnap generalize revision theory to work for all circular definitions, noting that the pathologies one sees with truth can be reproduced in many circular definitions as well. They make the following remark.

> Concepts with circular definitions, then, behave in ways that are remarkably similar to the behavior of the concept of truth. They exhibit the same kinds of pathological behavior as truth. And like truth, they can be, and usually are, unproblematic over a range of cases. . . . [T]he similarities suggest that the perplexing behavior of the concept of truth might be explainable as arising from some circularity in its definition.[8]

They move to circular definitions because they view the semantic paradoxes as arising due to some circularity in the definition of truth.

Circular definitions arise for Gupta and Belnap from reflection on and generalization from the case of truth. The set of Tarski biconditionals that defines the truth predicate constitutes a circular definition.

Let us turn to another route to circular definitions.

1.2 From meaning to circular definitions

Frege's distinction between sense and reference is well known. Recently, Yiannis Moschovakis has proposed a way to understand this distinction using algorithms and values.[9] In particular, senses are to be understood in terms of algorithms and reference in terms of values. Moschovakis's idea is, roughly

[8] Gupta and Belnap [1993, 117]

[9] See Moschovakis [1994, 2006].

this: The sense of a sentence is an algorithm that allows one to compute its truth value. Atomic formulas are interpreted as relations that can be immediately checked. To compute a truth-functional compound, one follows the algorithm for computing a truth function of the values of its parts. To compute the value of a quantified formula, one computes the value for each instance and then takes the maximum or minimum value.[10]

Sentences can be about themselves, and algorithms can call themselves without problem. Thus, circularity enters into this picture in a straightforward way. Indeed, Moschovakis cites sentences such as the liar,

(L): (L) is not true,

as motivation. One can understand the meaning of (L) in terms of an algorithm. To determine the semantic value of (L), one gets the value of (L) and returns 1 if the obtained value is 0, and 0 if the obtained value is 1. This algorithm loops without end, but there is no conceptual problem with an algorithm that does not terminate. Indeed, the value of (L) is $\frac{1}{2}$, here understood as meaning undefined.

Moschovakis formalizes these ideas using first-order logic extended with explicit self-reference devices. The explicit self-reference devices are new predicates, P_i, that are defined via formulas drawn from the extended language containing the new predicates. These new predicates have partial, i.e. three-valued, semantics. Their semantic values are fixed-points, which can be constructed in a way similar to the one indicated by Kripke.[11] Moschovakis is, then, led to a form of circular definitions that is formally equivalent to what will be presented below.[12]

Moschovakis's focus on computation leads him to focus on the Strong Kleene scheme.[13] As mentioned above, the third value is taken to mean undefined. He interprets the new predicates via fixed-points, focusing on the least fixed-points.[14] Although the formal aspects of the theory focus on the fixed-point interpretations, the philosophical motivations focus on algorithms

[10]Some computations will be infinite, in which case Moschovakis suggests using a broader notion of computation than standard ones.

[11]This is not surprising, since Kripke's construction was an instance of the inductive constructions of Moschovakis [2008].

[12]It is equivalent when only finitely many predicates are being defined. It is not clear that the equivalence extends to infinite sets of definitions.

[13]One might also be led to the Weak Kleene scheme on the basis of these considerations.

[14]Moschovakis [1994, 233] notes that the approach yields other fixed-points as well.

and procedures. For example, the value of (L) is $\frac{1}{2}$, but Moschovakis steps through the process of arriving at that, rather than stating that it has that value in all fixed-points. The focus on the algorithms and procedures suggests that the iterative constructions of fixed-points are philosophically important for Moschovakis. While the constructions do not play an important role in this paper, they are useful for understanding the semantics of circular definitions.

If one follows Moschovakis in viewing the sense of a sentence as an algorithm, then one has a natural motivation to study circular definitions. These will be algorithms that directly or indirectly call themselves and, in some cases, never return values.

Before proceeding to the immediate motivations for studying non-classical definitions, I will indicate a couple of potential payoffs of focusing on circular definitions.

1.3 Upshots

The preceding gave two historical routes philosophers have taken to arrive at circular definitions. Yet, one might wonder, if one doesn't take those routes, what is to be gained from circular definitions, as opposed to truth, properties, or sets.[15] There are four benefits that I will outline here.

First, circular definitions provide new options for analyzing interdependent concepts. One particular application for which the revision theory of circular concepts has been used is to develop an alternative account of rational choice in game theory.[16] While this conception of rationality agrees with the standard Nash equilibrium view in many cases, there are games that it evaluates differently and games that it can evaluate but that the Nash equilibrium view cannot.

Second, a logical upshot is the *classification* of definitions. In revision theory, certain classes of circular definitions naturally stand out. One prominent class is the class of *finite definitions*. This class is defined in terms of the behavior of the revision sequences involved; for finite definitions, roughly,

[15]There has been a lot of work on both truth and naive set theory recently. For the former, see Gupta and Belnap [1993], Field [2008], Beall [2009], Zardini [2011], Scharp [2013], and French [2016], among others. For the latter, see, for example, Restall [1992], Petersen [2000], Cantini [2003], Terui [2004], Weber [2010a], Weber [2010b], Weber [2012], Brady [2014], Omori [2015], and Ripley [2015], and references therein.

[16]See Chapuis [2003] and Gupta [2000].

nothing is gained from using transfinite sequences. Although there is no known syntactic specification of this class, its closure properties, axiomatizability, and complexity have been studied.[17]

Third is a philosophical upshot connected to theories of truth and paradox. As mentioned, Gupta and Belnap point out that circular definitions and truth are similar in many ways. If one follows Gupta and Belnap as taking the latter as a case of the former, then interest in truth should carry over to circular definitions, as the former is an instance of the latter. Suppose, on the contrary, that one does not follow Gupta and Belnap in taking truth to be circularly defined, as one might, for example, by denying that the Tarski biconditionals define truth. Even then, circular definitions can still appear to be paradoxical, in ways similar to truth and naive properties. If one thinks that solutions to the semantic paradoxes should be equally applicable to related concepts, such as sets, properties, and denotation, then it is natural to ask how how solutions to the paradoxes fair with respect to circular definitions. Circular definitions, then, can provide another area in which to compare different approaches to paradox.

Fourth, and finally, circular definitions can provide a different perspective on important or controversial principles in debates concerning theories of truth. As an example, I will briefly discuss one such principle: intersubstitutivity.

The Intersubstitutivity Principle (IP) says that A and $T(\ulcorner A \urcorner)$ are fully intersubstitutable in all extensional contexts, which is to say that if B differs from C only in having some occurrences of A, or $T(\ulcorner A \urcorner)$, replaced with $T(\ulcorner A \urcorner)$, or A, respectively, then B and C are equivalent. Fixed-point approaches to truth are often associated with (IP). Some philosophers, such as Hartry Field and Jc Beall, have claimed that truth must obey (IP) in order for it to fulfill its *generalizing function*.[18] An example of the generalizing function is in order. Suppose that Pr picks out the infinitely many provable sentences of an arithmetic theory with truth, with $\ulcorner 1 = 1 \urcorner$ among them. The generalizing function of truth allows one to go from

$$\forall x(Pr(x) \rightarrow T(x)),$$

to

$$Pr(\ulcorner 1 = 1 \urcorner) \rightarrow 1 = 1.$$

[17] See Martinez [2001] and Gupta [2006].

[18] See, respectively, Field [2008] and Beall [2009]. For a criticism of (IP), see Gupta and Standefer [2016].

This is an example of a situation that Quine commented on; he said that the truth predicate allows one to generalize on sentence position.[19] In so doing, the truth predicate permits one to affirm or deny possibly infinitely many sentences.

The main arguments in favor of (IP) are concerned with the logical behavior of the truth predicate, although related predicates, such as satisfaction and, sometimes, validity, are held as essentially obeying versions of (IP).[20] Rather than (IP), we can consider an analogous principle of intersubstitutivity for definitions. The arguments for intersubstitutivity face a stumbling block when adapted to circular definitions, since circular definitions generally do not have any sort of generalizing function to fulfill. An argument that circular definitions must obey an analog of (IP) will have to take another route. I will briefly look at three.

One possibility is to appeal to the Standard Theory of Definitions.[21] The Standard Theory says that it should be possible to eliminate a defined predicate from any context in which it appears. Of course, any theory of circular definitions cannot fully endorse that requirement, but it can endorse part of its spirit. While defined terms cannot always be eliminated, the theory would say that they can be replaced by their *definientia*, which may or may not contain the defined term.

Another possibility bypasses the Standard Theory and appeals to *meaning*. This view maintains that a *definiendum* and its *definiens* have the same meaning. They are intensionally equivalent, so substituting one for the other in extensional contexts should not change the truth value of the whole.

This point, however, is not obvious. There is a *prima facie* asymmetry in a definition between the *definiendum* and *definiens*. The former depends on the latter. This observation is not restricted to any particular theory. For example, Albert Visser, writing from a theory-neutral point of view, says, "Note that [the definitional connection $L =_{Df} \sim L$] is *prima facie* asymmetrical: the meaning of the right hand side is in some sense prior to the meaning of the left hand side."[22]

There is a third argument for intersubstitutivity for circular definitions that has nothing to do with definitions, per se, but rather with the underly-

[19] Quine [1986]

[20] See Shapiro [2011], Beall and Murzi [2013], and Zardini [2014] for discussions of a validity analog of (IP).

[21] See Belnap [1993] for discussion of the Standard Theory.

[22] Visser [2004, 168-169]

ing logic. If definitions are properly formalized using the biconditional of the logic, and that biconditional licenses a strong form of substitution of equivalents, then a version of intersubstitutivity for definitions will follow. Not all logics have biconditionals that license substitution of equivalents. Priest's LP is one for which the definitions would be formulated using a biconditional that does not validate substitution of equivalents.[23] Note that this point is dependent on the logical resources of the language, and will need amendment if, for example, modal operators are in the language.

It is important to note that the character of this argument for intersubstitutivity is rather different from the argument for (IP) in the context of truth. In the context of truth, (IP) is often taken as necessary to serve a generalizing role, facilitating quantification over sentences and expressing endorsement. That argument is not available for circular definitions, and there appears to be a question about whether and why one should try to salvage a version of intersubstitutivity for circular definitions.[24] As we will see, not all of the non-classical approaches to circular definitions obey intersubstitutivity for definitions. The Strong Kleene and LP theories do, but the supervaluation theory does not in general, putting it in the same camp as revision theory.

Let us turn from payoffs of circular definitions, to the more immediate motivations of this paper.

1.4 Motivations

I have sketched two ways into the study of circular definitions, as found in the work of Gupta and Belnap and the work of Moschovakis. This paper will follow in the spirit of Gupta and Belnap, rather than Moschovakis, primarily for two reasons. First, I do not adopt Moschovakis's philosophical motivation of understanding sense in terms of computation. That provides Moschovakis a reason to focus on the Strong Kleene scheme, but I do not want to restrict my attention to only that scheme. Second, I will focus on the *logic* of circular definitions. Given that the circular definitions will be interpreted in terms of fixed-points, the logic is naturally defined in terms of *all* models and *all* fixed-points. Moschovakis, by contrast, is more interested in *definability* in structures, which leads him to focus on particular fixed-points in particular models.

[23] I will return to the discussion of intersubstitutivity in §3.

[24] See Orilia and Varzi [1998] for consideration of rejecting intersubstitutivity to avoid the paradox of analysis.

I will study circular definitions in non-classical schemes, so I am, to an extent, departing from the *original* motivations for revision theory. I am following in the spirit of Gupta and Belnap in that they took an approach to truth and generalized it to circular definitions. In a similar vein, I will be looking at the theories of circular definitions one gets by moving from two non-classical approaches to truth to corresponding approaches to circular definitions. Something like this idea was suggested by Gupta and Belnap themselves, as they say, "The theoretical moves that have been made in response to the pathological behavior of truth can all be made with respect to circular concepts.... [T]he entire history of the Liar paradox can be mimicked in the context of circular definitions."[25]

This work is part of a larger project to compare non-classical approaches to circular definitions with the classical revision theories of circular definitions. I will explore circular definitions using the formal apparatus of fixed-point approaches to truth in two non-classical schemes.[26] There are many options for non-classical approaches to circular definitions, a point I return to in §5, but here I will focus on two three-valued schemes, with some asides about other three-valued schemes. I will not be able to provide a detailed comparison of revision theory with the two non-classical theories, since that would require more details on revision theory than I can provide here, but I will make some comparative comments along the way.

I will end this subsection by highlighting two features of revision theory that will guide some of the discussion to come: proof systems and defining clauses. First, the proof systems. The revision theory of circular definitions has a Fitch-style natural deduction proof system and a cut-free sequent system.[27] One of the important points of the subsequent sections is that the non-classical approaches have well-behaved proof systems, at least in the sense of being sound and complete, and so on that score are on equal footing with revision theory. The second feature concerns the defining clauses. In general, the quantified material biconditional versions of the defining clauses

[25] Gupta and Belnap [1993, 117]

[26] I will not be exploring definitions in revision theory using non-classical schemes. That project would be interesting, but it would take us too far afield.

[27] See Gupta and Belnap [1993, Ch. 5B] for more on the Fitch system. See Bruni [2013] for the sequent system. These proof systems are complete for a weak version of revision theory. Kremer [1993] has shown that a natural, strong version of revision theory is not axiomatizable.

of a set of circular definitions will not be valid in revision theory.[28] As we will see, not all of the non-classical approaches guarantee the validity of the quantified material biconditional versions of the defining clauses.

Now that I have presented two paths to thinking about circular definitions and some motivations for looking at particularly non-classical approaches, I will proceed to the non-classical definitions, focusing on Strong Kleene and supervaluation theories. Let us begin with Strong Kleene.

2 Strong Kleene definitions

For this section, I will consider circular definitions in the Strong Kleene scheme.[29] The Strong Kleene scheme has three semantic values, $1, 0, \frac{1}{2}$, which are linearly ordered by the logical ordering: $0 \leq_L \frac{1}{2} \leq_L 1$. The connectives for Strong Kleene are defined as follows.

$\sim$	
1	0
$\frac{1}{2}$	$\frac{1}{2}$
0	1

$\vee$	1	$\frac{1}{2}$	0
1	1	1	1
$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$
0	1	$\frac{1}{2}$	0

Let v be an assignment of values to variables. The universal quantifier has the semantic clause:

$$V_{M,v}(\forall x A(x)) = min(\{V_{M,v[d/x]}A(x) : d \in D\}),$$

where $v[d/x]$ is the assignment that v' that agrees with v on assignments to all variables apart from x, and $v'(x) = d$.[30] The other connectives and quantifier are defined out of these as usual.

As said at the outset, the base language is interpreted via a classical ground model $M(= \langle D, I \rangle)$. Hypotheses are functions from

[28] This does not mean that the the defining clauses fail to determine the meaning of the defined expressions. See Gupta and Belnap [1993, 253-255] for discussion. It has, however, recently been shown how to extend the revision theory with new biconditionals that validate versions of all the defining clauses; for details, see Standefer [2015] and Gupta and Standefer [2016].

[29] The following is based on the ccpo frameworks described in Gupta and Belnap [1993] and Visser [2004].

[30] I will generally suppress mention of assignments to variables and use the notation $A(\bar{d})$ to indicate evaluation under an assignment whose values on the free variables of A, $\bar{x}$, are the objects $\bar{d}$.

$\bigcup_n \{G : G \text{ is an } n\text{-ary } \textit{definiendum} \text{ of } \mathscr{D}\} \times D^n$ to $\{0, 1, \frac{1}{2}\}$. For all classical models M and hypotheses h, a model $M + h$ is just like M except that h is used to interpret *definienda* from $\mathscr{D}$, as follows for all n-ary *definienda* G and all $\bar{d} \in D^n$.

$$V_{M+h}(G(\bar{d})) = h(G)(\bar{d})$$

The semantic values are partially ordered by the information ordering: $\frac{1}{2} \leq_i 1$ and $\frac{1}{2} \leq_i 0$. This is used to define a partial order on the hypotheses of a given model.

Definition 1. *$h \preceq h'$ iff for all predicates G in $\mathscr{D}$ and all tuples $\bar{d}$ whose length is the arity of G, $h(G)(\bar{d}) \leq_i h'(G)(\bar{d})$.*

Each of these partial orders has a minimal element.

A set of definitions $\mathscr{D}$ yields a *jump* operator $\kappa_{\mathscr{D},M}$ that obeys the following constraint.

$$\kappa_{\mathscr{D},M}(h)(G)(\bar{d}) = V_{M+h}(A_G(\bar{d}))$$

The Strong Kleene scheme is monotonic in the sense that for all truth functions c, if $x_1 \leq_i y_1, \ldots, x_n \leq_i y_n$, then $c(x_1, \ldots, x_n) \leq_i c(y_1, \ldots, y_n)$.[31] It follows then that $\kappa_{\mathscr{D},M}$ is monotonic on $\preceq$, i.e.

$$h \preceq h' \Rightarrow \kappa_{\mathscr{D},M}(h) \preceq \kappa_{\mathscr{D},M}(h').$$

A hypothesis h is *sound* iff $h \preceq \kappa_{\mathscr{D},M}(h)$. A hypothesis f is a *fixed-point* iff $f = \kappa_{\mathscr{D},M}(f)$.

Given monotonicity, iterating $\kappa_{\mathscr{D},M}$, possibly transfinitely, on a sound h will yield a fixed-point, f with $h \preceq f$. For transfinite iterations, the limit stages are the joins of all previous stages.[32] In particular, iterating $\kappa_{\mathscr{D},M}$ on the $\preceq$-minimal hypothesis h_0 will yield the *minimal*, or least, fixed-point. The fixed-points f will be used to interpret defined predicates.

For a given language $\mathcal{L}$ and set of definitions $\mathscr{D}$, we can define Strong Kleene consequence, or entailment.

Definition 2 (Entailment)**.** *Let $\mathcal{L}$ be a base language and let $\mathscr{D}$ be a set of definitions. For all sentences $A_1, \ldots, A_n, B_1, \ldots, B_m$, $A_1, \ldots, A_n$ entails*

[31] All of the non-classical schemes discussed in this section and the next are monotonic.

[32] Not every set of of hypotheses will have a join, but since the hypotheses under consideration form are linearly ordered by $\preceq$, they do. Their join is the hypothesis that assigns 1, or 0, to a predicate and a tuple just in case some element of the chain assigns 1, or 0, respectively, to that predicate and tuple.

$B_1, \ldots, B_m$ in M on $\mathscr{D}$ (in symbols, $A_1, \ldots, A_n \models^{SK,M}_{\mathscr{D}} B_1, \ldots, B_m$) iff for all fixed-points f, if for all $i \leq n$, $V_{M+f}(A_i) = 1$, then for some $i \leq m$, $V_{M+f}(B_i) = 1$.

$A_1, \ldots, A_n$ entails $B_1, \ldots, B_m$ on $\mathscr{D}$ ($A_1, \ldots, A_n \models^{SK}_{\mathscr{D}} B_1, \ldots, B_m$) iff for all classical ground models M, $A_1, \ldots, A_n \models^{SK,M}_{\mathscr{D}} B_1, \ldots, B_m$.

A formula is $\mathscr{D}$-free if it contains no *definienda*. If all the A_i and B_j are $\mathscr{D}$-free, then entailment will reduce to classical entailment.

Strong Kleene entailment, for each $\mathscr{D}$, is axiomatized by the following sequent system, which is based on the sequent system for truth developed by Kremer [1988].[33] This system shares all its axioms and rules with Kremer's system except for the definition rules, which replace the truth rules of Kremer's system, and the contraction rules.

Axioms:

$$A \vdash^{SK}_{\mathscr{D}} A \qquad\qquad \vdash^{SK}_{\mathscr{D}} t = t$$

$$\sim A, A \vdash^{SK}_{\mathscr{D}}$$

$$\vdash^{SK}_{\mathscr{D}} B, \sim B \qquad\qquad t \neq t \vdash^{SK}_{\mathscr{D}}$$

In the axiom,

$$\vdash^{SK}_{\mathscr{D}} B, \sim B,$$

B must be $\mathscr{D}$-free.

Structural rules:

$$\frac{X \vdash^{SK}_{\mathscr{D}} Y}{A, X \vdash^{SK}_{\mathscr{D}} Y}\,(K\vdash) \qquad \frac{A, A, X \vdash^{SK}_{\mathscr{D}} Y}{A, X \vdash^{SK}_{\mathscr{D}} Y}\,(W\vdash)$$

$$\frac{X \vdash^{SK}_{\mathscr{D}} Y}{X \vdash^{SK}_{\mathscr{D}} Y, A}\,(\vdash K) \qquad \frac{X \vdash^{SK}_{\mathscr{D}} Y, A, A}{X \vdash^{SK}_{\mathscr{D}} Y, A}\,(\vdash W)$$

Connective rules:

[33] In Kremer's system, each side of the turnstile is a set, but here each side is a multiset.

$$\frac{X \vdash^{SK}_{\mathscr{D}} Y, A}{X \vdash^{SK}_{\mathscr{D}} Y, A \vee B} (\vdash \vee) \qquad \frac{\sim B, X \vdash^{SK}_{\mathscr{D}} Y}{\sim(A \vee B), X \vdash^{SK}_{\mathscr{D}} Y} (\sim\vee \vdash)$$

$$\frac{X \vdash^{SK}_{\mathscr{D}} Y, B}{X \vdash^{SK}_{\mathscr{D}} Y, A \vee B} (\vdash \vee) \qquad \frac{X \vdash^{SK}_{\mathscr{D}} Y, \sim A \qquad X \vdash^{SK}_{\mathscr{D}} Y, \sim B}{X \vdash^{SK}_{\mathscr{D}} Y, \sim(A \vee B)} (\vdash \sim\vee)$$

$$\frac{A, X \vdash^{SK}_{\mathscr{D}} Y \qquad B, X \vdash^{SK}_{\mathscr{D}} Y}{A \vee B, X \vdash^{SK}_{\mathscr{D}} Y} (\vee \vdash) \qquad \frac{A, X \vdash^{SK}_{\mathscr{D}} Y}{\sim\sim A, X \vdash^{SK}_{\mathscr{D}} Y} (\sim\sim \vdash)$$

$$\frac{\sim A, X \vdash^{SK}_{\mathscr{D}} Y}{\sim(A \vee B), X \vdash^{SK}_{\mathscr{D}} Y} (\sim\vee \vdash) \qquad \frac{X \vdash^{SK}_{\mathscr{D}} Y, A}{X \vdash^{SK}_{\mathscr{D}} Y, \sim\sim A} (\vdash \sim\sim)$$

Quantifier rules:

$$\frac{A[t/x], X \vdash^{SK}_{\mathscr{D}} Y}{\forall x A, X \vdash^{SK}_{\mathscr{D}} Y} (\forall \vdash) \qquad \frac{\sim A[y/x], X \vdash^{SK}_{\mathscr{D}} Y}{\sim\forall x A, X \vdash^{SK}_{\mathscr{D}} Y} (\sim\forall \vdash)$$

$$\frac{X \vdash^{SK}_{\mathscr{D}} Y, A[y/x]}{X \vdash^{SK}_{\mathscr{D}} Y, \forall x A} (\vdash \forall) \qquad \frac{X \vdash^{SK}_{\mathscr{D}} Y, \sim A[t/x]}{X \vdash^{SK}_{\mathscr{D}} Y, \sim\forall x A} (\vdash \sim\forall)$$

Identity rules:

$$\frac{X \vdash^{SK}_{\mathscr{D}} Y, A(s)}{s = t, X \vdash^{SK}_{\mathscr{D}} Y, A(t)} (\vdash =) \qquad \frac{X \vdash^{SK}_{\mathscr{D}} Y, A(s)}{X \vdash^{SK}_{\mathscr{D}} Y, A(t), s \neq t} (\vdash \neq)$$

$$\frac{A(s), X \vdash^{SK}_{\mathscr{D}} Y}{s = t, A(t), X \vdash^{SK}_{\mathscr{D}} Y} (= \vdash) \qquad \frac{A(s), X \vdash^{SK}_{\mathscr{D}} Y}{A(t), X \vdash^{SK}_{\mathscr{D}} Y, s \neq t} (\neq \vdash)$$

$$\frac{X \vdash^{SK}_{\mathscr{D}} Y, A(s)}{t = s, X \vdash^{SK}_{\mathscr{D}} Y, A(t)} (\vdash =) \qquad \frac{X \vdash^{SK}_{\mathscr{D}} Y, A(s)}{X \vdash^{SK}_{\mathscr{D}} Y, A(t), t \neq s} (\vdash \neq)$$

$$\frac{A(s), X \vdash^{SK}_{\mathscr{D}} Y}{t = s, A(t), X \vdash^{SK}_{\mathscr{D}} Y} (= \vdash) \qquad \frac{A(s), X \vdash^{SK}_{\mathscr{D}} Y}{A(t), X \vdash^{SK}_{\mathscr{D}} Y, t \neq s} (\neq \vdash)$$

In $(\vdash \forall)$ and $(\sim\forall \vdash)$, the variable y cannot occur freely in the conclusion

sequents. In the identity rules, t is free for s in A. The system also contains, for each definitional clause $G\bar{x} =_{Df} A_G(\bar{x})$ in $\mathscr{D}$, the following four rules.

$$\frac{A_G(\bar{t}), X \vdash^{SK}_{\mathscr{D}} Y}{G\bar{t}, X \vdash^{SK}_{\mathscr{D}} Y}\,(Def\vdash) \qquad \frac{\sim A_G(\bar{t}), X \vdash^{SK}_{\mathscr{D}} Y}{\sim G\bar{t}, X \vdash^{SK}_{\mathscr{D}} Y}\,(\sim Def\vdash)$$

$$\frac{X \vdash^{SK}_{\mathscr{D}} Y, A_G(\bar{t})}{X \vdash^{SK}_{\mathscr{D}} Y, G\bar{t}}\,(\vdash Def) \qquad \frac{X \vdash^{SK}_{\mathscr{D}} Y, \sim A_G(\bar{t})}{X \vdash^{SK}_{\mathscr{D}} Y, \sim G\bar{t}}\,(\vdash \sim Def)$$

For each set of definitions $\mathscr{D}$, call the sequent system resulting from the above rules $LSK(\mathscr{D})$. Then, $LSK(\mathscr{D})$ axiomatizes Strong Kleene entailment on $\mathscr{D}$.

Theorem 1. *$X \models^{SK}_{\mathscr{D}} Y$ iff $X \vdash^{SK}_{\mathscr{D}} Y$ is derivable in $LSK(\mathscr{D})$.*

The soundness direction is proved by a standard induction on the construction of a derivation. The completeness direction is proved by a Henkin-style argument, in the manner of the completeness proof of Kremer [1988]. Since there are no major changes to the proof, I will omit it.

The system $LSK(\mathscr{D})$ does not have cut,

$$\frac{X \vdash^{SK}_{\mathscr{D}} Y, A \qquad A, X' \vdash^{SK}_{\mathscr{D}} Y'}{X, X' \vdash^{SK}_{\mathscr{D}} Y, Y'}\,(cut)$$

among its basic rules.[34] The cut rule is sound, so one can use the soundness theorem with the premises of cut to obtain entailment analogs, which yield an entailment analog of the conclusion. Since $LSK(\mathscr{D})$ is complete, one can then infer that the conclusion of the cut rule is derivable, as desired. Thus, adding cut would result in no new sequents being provable. In other words, cut is admissible, so $LSK(\mathscr{D})$, without cut, is strong enough, from a proof-theoretic point of view. In this setting, however, cut admissibility does not have the usual consequences. In particular, for some $\mathscr{D}$, the subformula property can fail, which is to say that some derivations may contain predicates and connectives not appearing in their endsequents.[35] This can, to an extent, be remedied by defining an extended sense of subformula that takes as

[34] For each set of definitions $\mathscr{D}$, there is a sequent system $LSK(\mathscr{D})$. This paragraph should be read as applying to each of them.

[35] As usual, for quantified formulas we want to count instances as subformulas.

subformulas of a defined predicate all subformulas of its *definiens*, including subformulas of *definientia* of defined predicates occurring there, and so on. The rules preserve subformulas in this extended sense. Depending on the set of definitions $\mathscr{D}$, the extended sense of subformula may be informative or it may be trivial. If a set of definitions uses every predicate of the base language as well as all the logical connectives, then preserving extended subformulas would not say anything informative. If a set of definitions involves a relatively small fragment of the language, then it may be more informative. The usual subformula property, however, still holds when attention is restricted to derivations of sequents containing only base language predicates.

It is worth observing that entailment, as defined above, obeys the analog of (IP) for circular definitions,

$$A_1, \ldots, A_n \models^{SK}_{\mathscr{D}} B_1, \ldots, B_m \;\Leftrightarrow\; A'_1, \ldots, A'_n \models^{SK}_{\mathscr{D}} B'_1, \ldots, B'_m,$$

where C and C' differ at most by replacing, possibly zero, occurrences of *definientia* with their *definienda*, and vice versa, possibly relabeling bound variables. This follows straight away from the definition of entailment.

Just as the Tarski biconditionals are not all valid in the Strong Kleene theory of truth, there are definitions for which the (quantified) material biconditional versions of the definitional clauses are not valid. For example, the material biconditional version of the definition, $Lx =_{Df} {\sim}Lx$, will be invalid. I will return to this point after introducing the supervaluation theory of definitions, to which I now turn.

3 Supervaluation definitions

In this section, I will look at circular definitions in the supervaluation scheme.[36]

Say that a hypothesis h is *classical* just in case its range is $\{1, 0\}$. If h is classical, then let CL_{M+h} be the classical valuation of the model $M + h$. The supervaluation evaluation SV is defined as follows.

Definition 3.

$$SV_{M+h}(A) = \begin{cases} 1 & \textit{if } CL_{M+h'}(A) = 1, \textit{ for all classical } h' \succeq h \\ 0 & \textit{if } CL_{M+h'}(A) = 0, \textit{ for all classical } h' \succeq h \\ \frac{1}{2} & \textit{otherwise} \end{cases}$$

[36]The notation of this section follows that of Kremer and Urquhart [2008].

Definition 4 (Validity)**.** *Let $\mathcal{L}$ be a base language and let $\mathscr{D}$ be a set of definitions. For all sentences A of $\mathcal{L}^+$, A is valid on $\mathscr{D}$, $\models^{SV}_{\mathscr{D}} A$, iff for all ground models M, for all fixed-points f, $SV_{M+f}(A) = 1$.*

The supervaluation logic for $\mathscr{D}$ is $\{A \in Sent_{\mathcal{L}^+} : \models^{SV}_{\mathscr{D}} A\}$.

I have defined the supervaluation logic as a set of sentences, rather than a consequence relation, as for Strong Kleene definitions. This is motivated by results due to Kremer and Urquhart [2008], building on the work of Kremer and Kremer [2003], that suggest that the consequence relations for arbitrary definitions may be very complex.[37]

We can axiomatize the supervaluation logic for $\mathscr{D}$, following Kremer and Urquhart. Let the system $HSV(\mathscr{D})$ be the following.

- All classical, first-order logical truths.
- From A, infer any classical consequence of A.
- From $B \supset G(\bar{a})$, infer $B \supset A_G(\bar{a})$, where B is $\mathscr{D}$-free.
- From $B \supset {\sim}G(\bar{a})$, infer $B \supset {\sim}A_G(\bar{a})$, where B is $\mathscr{D}$-free.
- From $B \supset A_G(\bar{a})$, infer $B \supset G(\bar{a})$, where B is $\mathscr{D}$-free.
- From $B \supset {\sim}A_G(\bar{a})$, infer $B \supset G(\bar{a})$, where B is $\mathscr{D}$-free.

In the third and fourth rules, $\bar{a}$ must be free for $\bar{x}$ in $A_G(\bar{x})$. Define $\vdash^{SV}_{\mathscr{D}} A$ to mean that there is a proof of A in $HSV(\mathscr{D})$, with proof defined as usual for a Hilbert-style system.

Theorem 2. *Let $\mathcal{L}$ be a base language and let $\mathscr{D}$ be a set of definitions. Then, for all sentences A of $\mathcal{L}^+$,*

(1) If $\vdash^{SV}_{\mathscr{D}} A$, then $\models^{SV}_{\mathscr{D}} A$.

(2) If $\models^{SV}_{\mathscr{D}} A$, then $\vdash^{SV}_{\mathscr{D}} A$.

Part (1) is soundness and part (2) is completeness. The soundness proof proceeds by the usual induction on construction of derivations.

[37] The results just mentioned deal with truth, rather than circular definitions, but they are suggestive. Cf. Kremer [1993] on the complexity of definitions in revision theory.

The completeness proof proceeds much the same as that of Kremer and Urquhart [2008]. One starts by assuming $\nvdash^{SV}_{\mathscr{D}} A$. One then uses a Henkin-style construction to obtain a canonical, maximal, consistent fixed-point model which is, in general, three-valued and in which A does not receive semantic value 1. One then uses another Henkin-style construction to extend the three-valued model to a maximal, consistent classical model that assigns A the value 0. The classical model then serves as the hypothesis witnessing the fact that the canonical model does not assign A value 1. The primary difference between the proof for circular definitions and Kremer and Urquhart's original proof is a simplification: Kremer and Urquhart's proof, given for a language with a truth predicate, requires additional complications to account for the syntactic theory used for naming the sentences of the language. Since in this setting there is no syntactic theory that is added with the definitions, that aspect of the proof does not come in. The other major steps of the proof proceed here as there. Since there is substantial overlap, I omit the proof.

The supervaluation logic for $\mathscr{D}$ is defined for as a set of formulas. Define *logical consequence* for $\mathscr{D}$ as follows, where X is a set of sentences: $X \models^{SV}_{\mathscr{D}} B$ iff for all ground models M, for all fixed-points f over M, if $SV_{M+f}(A) = 1$, for all $A \in X$, then $SV_{M+f}(B) = 1$. It is plausible that for some, although not all, sets of definitions, consequence can be axiomatized in a way similar to that of the supervaluation logics. For example, let $\mathscr{D}$ be an ordinary definition, in the sense that no *definiens* contains any *definienda*. In that case, for each model, there is a unique fixed-point, and it is classical. It is plausible that consequence for such a $\mathscr{D}$ is axiomatizable. There is a question as to what conditions on definitions are sufficient to guarantee that the associated sense of logical consequence is axiomatizable.[38]

The system $HSV(\mathscr{D})$ is, arguably, natural, although it is not user-friendly, being a Hilbert-style axiom system. It would be good to have an alternative proof system, and I will provide one such here. I will define the Fitch-style natural deduction system $FSV(\mathscr{D})$ as follows. Let the rules for the connectives be any of the standard packages of complete rules for classical logic.[39]

[38]There are other ways of defining logical consequence in a supervaluation setting. The question holds for those as well.

[39]For example, the rules in Gupta and Belnap [1993, ch. 5B], minus the superscripts, will work.

To these rules, I add the following definition rules.

		$\vdots$				$\vdots$	
		$A_G(\bar{t})$				$\sim A_G(\bar{t})$	
		$G(\bar{t})$	DefIn			$\sim G(\bar{t})$	$\sim$DefIn
		$\vdots$				$\vdots$	
		$G(\bar{t})$				$\sim G(\bar{t})$	
		$A_G(\bar{t})$	DefElim			$\sim A_G(\bar{t})$	$\sim$DefElim

In these the last two rules, $\bar{t}$ must be free for $\bar{x}$ in $A_G(\bar{x})$. In all four of the definition rules, all of the assumptions in whose scope the displayed formulas appear must be $\mathscr{D}$-free. Define $\vdash^{FSV}_{\mathscr{D}} A$ to mean that A is derivable in $FSV(\mathscr{D})$ outside the scope of any assumptions. One can show that $\vdash^{SV}_{\mathscr{D}}$ and $\vdash^{FSV}_{\mathscr{D}}$ are coextensive.

Theorem 3. *Let $\mathscr{D}$ be a set of definitions. The following are equivalent for all sentences A.*

(1) $\vdash^{SV}_{\mathscr{D}} A$

(2) $\vdash^{FSV}_{\mathscr{D}} A$

Proof. For (2) to (1), it is clear that all the first-order logical truths are derivable in $FSV(D)$, and all classical logical consequences of a given formula will follow from the classical completeness of the underlying Fitch system. I will show that that the definition rules are also derivable by doing one case, the others being similar. I will show that if $B \supset A_G(\bar{t})$ is derivable, then $B \supset G(\bar{t})$ is derivable, where B is $\mathscr{D}$-free. Assume $B \supset A_G(\bar{t})$ is derivable. The $FSV(\mathscr{D})$ proof then looks like the following.

1	$B \supset A_G(\bar{t})$	
2	B	Hyp
3	$B \supset A_G(\bar{t})$	Reit 1
4	$A_G(\bar{t})$	$\supset$Elim 2, 3
5	$G(\bar{t})$	DefIn 4
6	$B \supset G(\bar{t})$	$\supset$In 2–5

Note that the DefIn rule at line 5 is justified because $A_G(\bar{t})$ is only within the scope of B, which is, by assumption, $\mathscr{D}$-free.

For (1) to (2), the proof is a variation on the usual proof showing Hilbert and Fitch systems equivalent.[40] Given a Fitch-style proof with no undischarged assumptions, we show how to generate a Hilbert-style proof of the same conclusion. First, we inductively turn each innermost subproof into a sequence of conditionals whose antecedents are the assumption of the subproof, prefixing the conditionals with universal quantifiers in the case of quantified subproofs. The result will be a sequence of conditionals that are derivable in $FSV(\mathscr{D})$, but which sequence may not form a $FSV(\mathscr{D})$ proof. We then show that each conditional in this sequence is derivable in $HSV(\mathscr{D})$.

Each conditional in the sequence corresponds, in a straightforward way, with a step of the Fitch proof, which has an associated rule justifying it. For the steps labelled with rules of classical logic, it is clear that one can derive the corresponding conditionals in $HSV(\mathscr{D})$ using classical axioms and rules from what came before. It remains to show that the definition rules can be obtained. I will do one case, DefIn. There are two subcases. In case $A_G(\bar{t})$ appears outside the scope of any assumptions in the original Fitch-style proof, then the corresponding "conditional" is simply $A_G(\bar{t})$, which implies $\top \supset A_G(\bar{t})$, whence one derives $G(\bar{t})$ from a definition rule and *modus ponens*. Otherwise, the conditional for this step has the form $B_1 \supset (B_2 \supset \ldots (B_n \supset A_G(\bar{t}))\ldots)$. This implies $(B_1 \,\&\, \ldots \,\&\, B_n) \supset A_G(\bar{t})$. Since all of the B_i are $\mathscr{D}$-free, by assumption, we can apply the definition rule in $HSV(\mathscr{D})$ to obtain $(B_1 \,\&\, \ldots \,\&\, B_n) \supset G(\bar{t})$, whence we obtain the desired $B_1 \supset (B_2 \supset \ldots (B_n \supset G(\bar{t}))\ldots)$. The cases for the other definition rules are similar. □

Gupta [2006] asks whether there is a natural sound and complete calculus for finite definitions in the supervaluation scheme. This section shows that there is an arguably natural calculus, $HSV(\mathscr{D})$. I have offered an improvement, albeit modest, on this in the form of $FSV(\mathscr{D})$, which I think answers Gupta's question in the affirmative. It would be good to have additional alternative proof systems, and I will briefly mention two possible directions to pursue to this end. Recently, Welch [2009] presented games for determining membership in the least supervaluation fixed-point for truth over the standard model of arithmetic, and Meadows [2015] presented a tableau system for membership in that fixed-point. It seems promising to use their ideas to develop a tableau system for supervaluation validity.

[40] See Anderson and Belnap [1975] for examples of this method of proof.

The previous section ended with a look at intersubstitutivity and definitional equivalences, and I will close this section in the same way. In contrast to the Strong Kleene consequence relation, the supervaluation logic does not obey the intersubstitutability principle. There are definitions for which $C(G\bar{t})$ is valid but $C(A_G(\bar{t}))$ is not. For example, let $\mathscr{D}$ be $Lx =_{Df} {\sim}Lx$. Then, $La \vee {\sim}La$ is valid while ${\sim}La \vee {\sim}La$ is not. Equivalence between *definiens* and *definiendum* holds, however, in freestanding contexts, so $G\bar{t}$ is valid iff $A_G(\bar{t})$ is valid and ${\sim}G\bar{t}$ is valid iff ${\sim}A_G(\bar{t})$ is valid.

It is also worth observing that, as with the Strong Kleene theory of definitions, not all (quantified) material biconditional versions of the defining clauses will be valid. For the definition of the previous paragraph, the corresponding material biconditional would be $\forall x(Lx \equiv {\sim}Lx)$, which will be contravalid, as it is false in all classical hypotheses. The supervaluation and Strong Kleene theories of definitions will differ on the validity of the material biconditional versions of some definitions. For example, the defining clause $Hx =_{Df} Hx$ will have as its material counterpart $\forall x(Hx \equiv Hx)$, which is valid in the supervaluation theory but not in the Strong Kleene theory.

Before turning to the discussion of a particular class of definitions, I will remark on alternative schemes. The schemes on which I have focused have natural duals, namely the LP and subvaluation schemes.[41] These schemes evaluate formulas in the same way as the Strong Kleene and supervaluation schemes, respectively, but they take both 1 and $\frac{1}{2}$ to be designated. The proof system of §2 can easily be adapted to LP. It is an open question whether one can obtain an axiomatization of the subvaluation logic for $\mathscr{D}$ by any straightforward adaptation of $HSV(\mathscr{D})$.

There is a feature of the LP and subvaluation theories of circular definitions to which I want to draw attention. As mentioned above, the material biconditional versions of the defining clauses are not always valid in the Strong Kleene and supervaluation theories. They are, however, always valid in the LP theory. For some definitions, the negations of the biconditionals are also valid, as one might expect. The subvaluation theory, by contrast, shares with the supervaluation theory, as well as classical revision theory, the feature of validating only the negation of some of the material biconditionals. On this front, then, the LP theory has a striking advantage over the other

[41] For more on LP, see, for example, Priest [1979] or Priest [2008]. For more on the subvaluation scheme, see Hyde [1997] or Cobreros [2013]. I would like to thank Graham Priest for pointing me towards the subvaluation scheme.

fixed-point theories mentioned: The material biconditionals corresponding to the definitional equivalences are always valid. Although it is not a decisive point, the lack of such valid biconditionals does leave a question hanging over the theories, namely, how can one express the relevant sense of definitional connection? This question is, I think, most pressing for the theories in which the negations of some of the biconditionals are valid.

Let us now turn from definitions in general to a particular class of them.

4 Intrinsic definitions

As mentioned in §1.3, different theories of definitions will highlight different classes of definitions. One class that emerges naturally in the Strong Kleene and supervaluation theories is the class of *intrinsic definitions*. The idea, based on an idea from Kripke [1975], is that hypotheses for intrinsic definitions do not make any arbitrary assignments. Many definitions are not intrinsic, in the sense to be defined. In the four-valued FDE scheme, by contrast, every definition is intrinsic. For the remainder of the paper, I will focus on the intrinsic definitions.

Definition 5. *Given language $\mathcal{L}$, a set of definitions $\mathscr{D}$, and a model M, let a fixed-point h be* κ-intrinsic*, or σ-intrinsic, just in case for all fixed-points h', there is a fixed-point h'' such that $h \preceq h''$ and $h' \preceq h''$, in the Strong Kleene, respectively supervaluation, scheme.*

Let a definition $\mathscr{D}$ be κ-intrinsic*, or σ-intrinsic, iff for all models, all fixed-points for $\mathscr{D}$ are κ-intrinsic or σ-intrinsic, respectively.*

The κ- and σ-intrinsic definitions are naturally isolated classes of definitions in the Strong Kleene and supervaluation theories of definitions, respectively. Every definition, intrinsic or not, will have a greatest intrinsic fixed-point, although in some cases it may simply be the minimal fixed-point.[42] Some examples are in order. Two simple κ- and σ-intrinsic definitions are

$$Lx =_{Df} {\sim}Lx$$

and

$$Hx =_{Df} Hx \vee {\sim}Hx.$$

[42] See Gupta and Belnap [1993, 69] for a proof. Gupta and Belnap attribute the point to Kripke [1975].

On these definitions, in both the Strong Kleene and supervaluation schemes, sentences of the form La can only the value $\frac{1}{2}$ in any fixed-point, and sentences of the form Ha cannot take the value 0 in any fixed-point. An example of a definition that is not intrinsic is

$$Kx =_{Df} Kx.$$

Suppose that $a \in D$; then there are fixed-points f and g for this definition such that $f(K)(a) = 1$ and $g(K)(a) = 0$, which suffices for f and g to be incomparable with respect to $\preceq$.

Intrinsic definitions in the two schemes are not identical, as there are definitions that are κ-intrinsic but not σ-intrinsic. Here is an example.[43] Let $\mathscr{D}$ contain just

$$Gx =_{Df} {\sim}Gx \text{ and } Hx =_{Df} Hx \vee (Gx \,\&\, {\sim}Gx),$$

and consider a model M. The Strong Kleene fixed-points are all intrinsic, since they can assign only $\frac{1}{2}$ to the G parts of H, which makes it impossible to have $f(H)(a) = 0$. On the other hand, $CL_{M+h}(Ga \,\&\, {\sim}Ga) = 0$, for all classical hypotheses h, so there will be fixed-points f and g such that $f(H)(a) = 1$ and $g(H)(a) = 0$.

The intrinsic definitions arise naturally when considering features of fixed-points. There is another feature of intrinsic definitions that is worth mention, especially for comparison with revision theory. For this brief comparison, I will focus on the Strong Kleene fixed-point approach and κ-intrinsic definitions. The κ-intrinsic definitions can be split into several sets, but there are two of particular interest. One set has, roughly speaking, few fixed-points per model. In some cases, this results in the definition's fixed-points assigning only $\frac{1}{2}$ to tuples. In other cases, it results in unique fixed-points, which are classical. The other set has, intuitively, many fixed-points, which has the result that its consequence relation is, in a sense, highly discriminating for atoms containing the defined predicate. Rather than going into the details of the two sets here, I will give some examples. I will assume that there are countably many names in the language, a and b_i, $i \geq 0$.

Example 1. *Let $\mathscr{L}$ be $Lx =_{Df} {\sim}Lx$. $\mathscr{L}$ is κ-intrinsic, and it has only one fixed-point f per model M, namely $f(G)(d) = \frac{1}{2}$, for each $d \in D$. We have the following.*

[43] I owe this example to Anil Gupta.

- $a \neq b_i, La \models^{SK}_{\mathscr{L}} Lb_i$.
- *For all sentences* B, $La \models^{SK}_{\mathscr{L}} B$.

Example 2. *Let* $\mathscr{H}$ *be* $Hx =_{Df} Hx \vee {\sim}Hx$. *Again,* $\mathscr{H}$ *is* κ*-intrinsic, but this time it has at least 2 fixed-points per model. For every* $a \in D$, *there are fixed-points* f, g, *with* $f(G)(a) = \frac{1}{2}$ *and* $g(G)(a) = 1$ *and* $f(G)(b) = g(G)(b)$, *for* $b \neq a$. *We then have the following.*

- $a \neq b_i, Ha \not\models^{SK}_{\mathscr{H}} Hb$.
- *It is not the case that for all sentences* B, $Ha \models^{SK}_{\mathscr{H}} B$.

What is the connection to revision theory? Although there are some caveats, revision theory and the Strong Kleene fixed-point theory take roughly dual stances on these definitions and definitions similar to them. Take $\mathscr{L}$ from the first example above. The consequence relation for this definition, as given by revision theory, will behave like that in the second example. By contrast, the revision-theoretic consequence relation for $\mathscr{H}$ will behave like that in the first example, at least with respect to atoms involving H.[44] The reason is that when there are not many values for the Strong Kleene semantics to assign to atoms with defined predicates, there are many values in the revision semantics, but the Strong Kleene semantics can assign a range of values in cases in which revision semantics can assign only one. Rather than pursue this point, I will leave it and turn to some features of intrinsic definitions.

There is a simple sufficient condition for a definition being κ- or σ-intrinsic. Say that a sentence B has the form of a classical validity, or contradiction, if it is a substitution instance of a classical, first-order validity, or contradiction, respectively.

Lemma 1. *If a set of definitions* $\mathscr{D}$ *is such that each* definiens *has the form of a classical logical validity, then there is no model* M *and fixed-point* f *such that* $V_{M+f}(G(\bar{a})) = 0$, *or* $SV_{M+f}(G(\bar{a})) = 0$.

Theorem 4. *If a set of definitions* $\mathscr{D}$ *is such that each* definiens *has the form of a classical logical validity, then* $\mathscr{D}$ *is both* κ- *and* σ*-intrinsic.*

[44] The consequence relation will, however, lack the second feature, namely Ha entailing B, for all sentences B.

Proof. Suppose not, so that each *definiens* of $\mathscr{D}$ has the specified form but that there are two incomparable fixed-points. So, there is a model M, fixed-points f and g, and a tuple $\bar{a}$ from M such that $f(G)(\bar{a}) = 1$ while $g(G)(\bar{a}) = 0$. Since $A_G(\bar{a})$ has the form of a classical validity, by the previous lemma, $V_{M+g}(A_G(\bar{a}))$, and $SV_{M+g}(A_G(\bar{a}))$, cannot be 0. But, M, f, g, and $\bar{a}$ were the witnesses to the assumed existence of incomparable fixed-points. Therefore, $\mathscr{D}$ is intrinsic. □

This sufficient condition is not necessary. It admits dualization, since if each *definiens* has the form of a classical contradiction, then the definition will also be intrinsic. One can also mix and match for definitions with multiple clauses. The condition readily admits generalization and refinement, some of which I will present now.

The propositions shown for Strong Kleene definitions involve valid or contradictory forms, which end up being simply 1 or 0, respectively, in the supervaluation scheme. An example of a definition that is σ-intrinsic and that can take the value $\frac{1}{2}$ is the following.

$$Gx =_{Df} Hx \quad Hx =_{Df} Hx \vee {\sim}Gx$$

The fixed-points cannot assign 0, and each object receives the same value for both G and H.

Say that a definition is *pure* just in case it contains no base language predicates, including identity, function symbols, or names. One can use tableau methods to settle whether a pure definition is κ-intrinsic. The rules for the tableaux are those for classical logic, together with the following rules.[45]

- Extend a branch containing a node $G\bar{t}$ with $A_G(\bar{t})$, and
- extend a branch containing a node ${\sim}G\bar{t}$ with ${\sim}A_G(\bar{t})$.

As usual, a tableau is complete if all rules that can be applied have been applied. A branch in the tableau is closed if it contains both B and ${\sim}B$, for any formula B. A tableau is closed if all branches are closed. Define an intrinsic tableau to be a complete tableau generated by the above rules.

[45] For a presentation of classical tableaux, along with proofs of standard metatheoretic results, see Smullyan [1995] or Priest [2008], among others.

Lemma 2. *Let $\mathscr{D}$ be a pure definition containing only the definitional clause, $G\bar{x} =_{Df} A_G(\bar{x})$. Let M be a ground model and let f be a fixed-point. If the premiss nodes in an intrinsic tableau get the value 1 in $M + f$, then at least one of the conclusion nodes does.*

Proof. The proof is by induction on the construction of the tableau. The classical rules are immediate from the standard soundness proofs. The cases covering the two definition rules are immediate from the definition of fixed-point. □

Proposition 1. *Let $\mathscr{D}$ be a pure definition with one definitional clause, $G\bar{x} =_{Df} A_G(\bar{x})$. Add new names $\bar{a}$ to the language. If the intrinsic tableau for either $\sim G\bar{a}$ or $G\bar{a}$ closes, then $\mathscr{D}$ is κ-intrinsic.*

Proof. Suppose that the intrinsic tableau for $\sim G\bar{a}$ closes but $\mathscr{D}$ is not κ-intrinsic. Then there is a model M and there are fixed-points f and g such that $f(G)(\bar{a}) = 1$ and $g(G)(\bar{a}) = 0$. Then $M + g$ is a model assigning the top node of the tableau the semantic value 1. By the previous lemma, there is a branch in which every node gets the value 1. But, that branch closes, so there is a formula B such that B and $\sim B$ receive 1, which is impossible. Therefore $\mathscr{D}$ is κ-intrinsic.

The case in which $G\bar{a}$ closes is similar. □

Following Martinez [2001], it is natural to ask whether the intrinsic definitions are closed under any syntactic operations. The results are primarily negative. For the rest of this section, I will not restrict attention to the pure definitions.

To start, I need to define some notation. If G and H are both n-ary, let the formula $B[H/G]$ be the result of replacing all occurrences of G in B with H but keeping the argument of each occurrence of H. For example, $G\bar{t}[H/G]$ is $H\bar{t}$. If a *definiens* A_G is under consideration, I will use the notation $A_{[H/G]}$ to mean $A_G[H/G]$.

Define the negation of a definition $G\bar{x} =_{Df} A_G(\bar{x})$ to be $H\bar{x} =_{Df} \sim A_{[H/G]}$

Proposition 2. *The κ-intrinsic definitions are not closed under negation.*

Proof. The definition $Lx =_{Df} \sim Lx$ is κ-intrinsic but $Mx =_{Df} \sim\sim Mx$ is not. □

Two operations under which the set of intrinsic definitions is not closed are *modus ponens* and self-composition.[46] Let $\mathscr{D}$ be $G\bar{x}=_{Df}A_G(\bar{x})$ and let $\mathscr{D}'$ be $G\bar{x}=_{Df}B_G(\bar{x})$, with both Gs n-ary. Let $\mathscr{D}\supset\mathscr{D}'$ be $G\bar{x}=_{Df}A_G(\bar{x})\supset B_G(\bar{x})$. Say that a class of one clause definitions is closed under *modus ponens* just in case if $\mathscr{D}$ and $\mathscr{D}\supset\mathscr{D}'$ are both in the class, then so is $\mathscr{D}'$.

Proposition 3. *The κ-intrinsic definitions with one clause are not closed under* modus ponens.

Proof. I show that $\mathscr{D}$ and $\mathscr{D}\supset\mathscr{D}'$ can both be κ-intrinsic without $\mathscr{D}'$ being so. Let $\mathscr{D}$ be $Gx=_{Df}x\neq x$ and let $\mathscr{D}'$ be $Gx=_{Df}Gx$. The definition $\mathscr{D}\supset\mathscr{D}'$ is $Gx=_{Df}x\neq x\supset Gx$. Clearly, $\mathscr{D}$ is intrinsic, as is $\mathscr{D}\supset\mathscr{D}'$. $\mathscr{D}'$, however, is not. □

Define the n-fold self-composition $\mathscr{D}^n$ of a definition $\mathscr{D}$ $G\bar{x}=_{Df}A_G(\bar{x})$ as follows.

- $A^0_G(\bar{x})=G\bar{x}$
- $A^{n+1}_G(\bar{x},G)=A^n_G(\bar{x})[A_G(\bar{t})/G\bar{t}]$, where the right-hand side is obtained by substituting $A_G(\bar{t})$ for each occurrence of $G\bar{t}$ in $A^n_G(\bar{x})$, relettering bound variables if needed.

Say that a class of definitions is closed under self-composition iff if $\mathscr{D}$ is in the class, then $\mathscr{D}^n$ is in the class, for all $n\geq 1$.

Proposition 4. *The κ-intrinsic definitions are not closed under self-composition.*

Proof. The definition $\mathscr{L}$ $Lx=_{Df}\sim Lx$ is κ-intrinsic, but $\mathscr{L}^2$ is $Lx=_{Df}\sim\sim Lx$, which is not intrinsic. □

Let $\mathscr{D}$ contain only $G\bar{x}=_{Df}A_G(\bar{x})$ and let $\mathscr{E}$ contain only $H\bar{x}=_{Df}B_H(\bar{x})$, with both defined predicates n-ary. Let the *conjunction* of the definitions be $J\bar{x}=_{Df}A_{[J/G]}(\bar{x})\,\&\,B_{[J/H]}(\bar{x})$. Similarly, if $\mathscr{D}$ contain only $G\bar{x}=_{Df}A_G(\bar{x})$ and $\mathscr{E}$ contain only $H\bar{x}=_{Df}B_H(\bar{x})$, then say their *disjunction* is $J\bar{x}=_{Df}A_{[J/G]}(\bar{x})\vee B_{[J/H]}(\bar{x})$. Finally, given a definition $\mathscr{D}$ containing only $G(\bar{x},y)=_{Df}A_G(\bar{x},y)$, its *universal quantification* is $J\bar{x}=_{Df}\forall yA_{[J/G]}(\bar{x})$. It is an open question whether the n-ary κ-intrinsic definitions are closed under these operations.

[46] These operations are studied in the context of the finite definitions of revision theory by Martinez [2001].

There is a related, but trivial, sort of closure to distinguish from closure under conjunction, disjunction, and universal quantification. Say that a definitional abstraction of a formula $A(\bar{x})$ is a defining clause $G\bar{x} =_{Df} A(\bar{x})$ for a new predicate, G. Closure under conjunction, disjunction, and quantifiers should not be confused with the trivial closures under abstractions. Given two definitions, $G\bar{x} =_{Df} A_G(\bar{x})$ and $H\bar{x} =_{Df} B_H(\bar{x})$, construct a new definition containing the previous two clauses and $J\bar{x} =_{Df} A_G(\bar{x}) \,\&\, B_H(\bar{x})$, which is a definitional abstraction. The fixed-points for J can be determined in terms of the fixed-points for G and H. The fixed-points for the conjunction of two definitions, by contrast, does not appear to be straightforwardly definable in terms of the fixed-points of the conjoined definitions.

I have focused on the κ-intrinsic definitions, which exhibit many of the same failures of closure that the finite definitions of revision theory do. The κ-intrinsic definitions are not, however, closed under self-composition. The question of whether the κ-intrinsic definitions are closed under some other plausible closure conditions, such as closure under conjunction, has been left open. I did not investigate LP-intrinsic definitions explicitly, because there is no need. The definition of intrinsicality appealed only to fixed-points and their ordering, which in turn depend only on the evaluation of formulas and the definitions involved. Since LP and Strong Kleene use the same evaluations for their connectives, fixed-points in the two schemes are identical; their differences emerge at the level of consequence. I have not, however, touched the closure properties of the σ-intrinsic definitions.

5 Conclusions

In this paper, I motivated the study of circular definitions and presented the basics of two different theories of circular definitions, the Strong Kleene and supervaluation theories. In the context of those theories, I identified and explored a particular class of definitions, the intrinsic definitions. Even considering just three-valued schemes, there is still more to explore. As I indicated in §3, the LP and subvaluation theories provide interesting contrasts to the Strong Kleene and supervaluation theories. Of the theories mentioned, only the LP theory validates all the material biconditional versions of the definitional clauses. A more detailed comparison between classical revision theory and theories in three-valued schemes would be illuminating, although

I will not pursue that here.[47]

There are, of course, other non-classical logics that could be used for circular definitions. I will mention two that seem particularly worth investigation, both of which involve relevant conditionals. First is the logic defended by Brady [2006], which is a depth relevance logic based on the concept of meaning containment. One could take the definitional clauses to be quantified biconditionals in this logic and then add them as object language axioms to theories. A plausible philosophical picture emerges: One takes the meaning of the *definiendum* to be identical to that of the *definiens*, as expressed by the object language biconditional. The second is the relevant logic R of Anderson and Belnap [1975].[48] It is well known that adding all the Tarski biconditionals or all of the set comprehension axioms to R results in triviality.[49] Suppose one takes definitional clauses to be quantified R biconditionals. Some definitions will be trivial, e.g. $Cx =_{Df} Cx \to \bot$. Others, however, would not be, including $Lx =_{Df} {\sim}Lx$ and $Kx =_{Df} Kx \to p$, the latter of which might be odd, as the particular p would be derivable, but still non-trivial. One does not have to use all circular definitions at once, so one could use circular definitions to analyze some concepts without necessarily running into the trivializing paradoxes of the full comprehension scheme of naive set theory.

Acknowledgments

I would like to thank Anil Gupta, Greg Restall, an anonymous referee, and the audiences at the Munich Center for Mathematical Philosophy, the Melbourne Logic Seminar, and the Frontiers of Non-Classicality Conference. This paper has benefitted greatly from their feedback.

This research was supported by the Australian Research Council, Discovery Grant DP150103801.

[47] The non-transitive logic for truth proposed by Cobreros et al. [2013] and Ripley [2013] would round out the comparison.

[48] Similar points will hold for Anderson and Belnap's logic E of entailment.

[49] See Meyer et al. [1979], for example.

References

Alan Ross Anderson and Nuel D. Belnap, Jr. *Entailment: The Logic of Relevance and Necessity*, volume 1. Princeton University Press, 1975.

Jc Beall. *Spandrels of Truth*. Oxford University Press, 2009.

Jc Beall and Julien Murzi. Two flavors of Curry's paradox. *Journal of Philosophy*, 110(3):143–165, 2013. doi:10.5840/jphil2013110336.

Nuel Belnap. On rigorous definitions. *Philosophical Studies*, 72(2/3):115–146, 1993. doi:10.1007/BF00989671.

Nuel D. Belnap, Jr. Gupta's rule of revision theory of truth. *Journal of Philosophical Logic*, 11(1):103–116, 1982. doi:10.1007/BF00302340.

Ross Brady. *Universal logic*. CSLI Publications, 2006.

Ross T. Brady. The consistency of the axioms of abstraction and extensionality in three-valued logic. *Notre Dame Journal of Formal Logic*, 12: 447–453, 1971. doi:10.1305/ndjfl/1093894366.

Ross T. Brady. The simple consistency of naive set theory using metavaluations. *Journal of Philosophical Logic*, 43(2-3):261–281, 2014. doi:10.1007/s10992-012-9262-2.

Riccardo Bruni. Analytic calculi for circular concepts by finite revision. *Studia Logica*, 101(5):915–932, 2013. doi:10.1007/s11225-012-9402-2.

Andrea Cantini. The undecidability of Grišin's set theory. *Studia Logica*, 74 (3):345–368, 2003. doi:10.1023/A:1025159016268.

André Chapuis. An application of circular definitions: Rational decision. In Benedikt Löwe, Wolfgang Malzkorn, and Thoralf Räsch, editors, *Foundations of the Formal Sciences II*, pages 47–54. Kluwer, 2003.

Pablo Cobreros. Vagueness: Subvaluationism. *Philosophy Compass*, 8(5): 472–485, 2013. doi:10.1111/phc3.12030.

Pablo Cobreros, Paul Égré, Robert van Rooij, and David Ripley. Reaching transparent truth. *Mind*, 122(488):841–866, 2013. doi:10.1093/mind/fzt110.

Hartry Field. *Saving Truth From Paradox*. Oxford University Press, 2008.

Rohan French. Structural reflexivity and the paradoxes of self-reference. *Ergo, an Open Access Journal of Philosophy*, 3, 2016. doi:10.3998/ergo.12405314.0003.005.

Paul C. Gilmore. The consistency of partial set theory without extensionality. In Thomas Jech, editor, *Axiomatic Set Theory*, volume 13 of *Proceedings of Symposia in Pure Mathematics*. American Mathematical Society, 1974.

Anil Gupta. Truth and paradox. *Journal of Philosophical Logic*, 11(1):1–60, 1982. doi:10.1007/BF00302338.

Anil Gupta. On circular concepts. In André Chapuis and Anil Gupta, editors, *Circularity, Definition and Truth*, pages 123–153. Indian Council of Philosophical Research, 2000. Reprinted in Gupta [2011], pp. 95-134.

Anil Gupta. Finite circular definitions. In Thomas Bolander, Vincent F. Hendricks, and Stig Andur Andersen, editors, *Self-Reference*, pages 79–93. CSLI Publications, 2006.

Anil Gupta. *Truth, Meaning, Experience*. Oxford University Press, 2011.

Anil Gupta and Nuel Belnap. *The Revision Theory of Truth*. MIT Press, 1993.

Anil Gupta and Shawn Standefer. Conditionals in theories of truth. *Journal of Philosophical Logic*, pages 1–37, 2016. doi:10.1007/s10992-015-9393-3. Forthcoming.

Hans G. Herzberger. Notes on naive semantics. *Journal of Philosophical Logic*, 11(1):61–102, 1982. doi:10.1007/BF00302339.

Dominic Hyde. From heaps and gaps to heaps of gluts. *Mind*, 106(424): 641–660, 1997. doi:10.1093/mind/106.424.641.

Michael Kremer. Kripke and the logic of truth. *Journal of Philosophical Logic*, 17(3):225–278, 1988. doi:10.1007/BF00247954.

Philip Kremer. The Gupta-Belnap systems $S^{\#}$ and S^{*} are not axiomatisable. *Notre Dame Journal of Formal Logic*, 34(4):583–596, 1993. doi:10.1305/ndjfl/1093633907.

Philip Kremer and Michael Kremer. Some supervaluation-based consequence relations. *Journal of Philosophical Logic*, 32(3):225–244, 2003. doi:10.1023/A:1024240422978.

Philip Kremer and Alasdair Urquhart. Supervaluation fixed-point logics of truth. *Journal of Philosophical Logic*, 37(5):407–440, 2008. doi:10.1007/s10992-007-9071-1.

Saul Kripke. Outline of a theory of truth. *Journal of Philosophy*, 72(19): 690–716, 1975. doi:10.2307/2024634.

Robert L. Martin and Peter W. Woodruff. On representing 'true-in-L' in L. *Philosophia*, 5(3):213–217, 1975. doi:10.1007/BF02379018.

Maricarmen Martinez. Some closure properties of finite definitions. *Studia Logica*, 68(1):43–68, 2001. doi:10.1023/A:1011998021743.

Toby Meadows. Infinitary tableau for semantic truth. *Review of Symbolic Logic*, 8(2):207–235, 2015. doi:10.1017/S175502031500012X.

Robert K. Meyer, Richard Routley, and J. Michael Dunn. Curry's paradox. *Analysis*, 39(3):124–128, 1979. doi:10.1093/analys/39.3.124.

Yiannis N. Moschovakis. Sense and denotation as algorithm and value. In J. Oikkonen and J. Väänänen, editors, *Lecture Notes in Logic*, number 2, pages 210–249. Springer, 1994.

Yiannis N. Moschovakis. A logical calculus of meaning and synonymy. *Linguistics and Philosophy*, 29(1):27–89, 2006. doi:10.1007/s10988-005-6920-7.

Yiannis N. Moschovakis. *Elementary Induction on Abstract Structures*. Dover Publications, 2008.

Hitoshi Omori. Remarks on naive set theory based on LP. *Review of Symbolic Logic*, 8(2):279–295, 2015. doi:10.1017/S1755020314000525.

Francesco Orilia and Achille C. Varzi. A note on analysis and circular definitions. *Grazer Philosophische Studien*, 54:107–113, 1998. doi:10.5840/gps19985428.

Uwe Petersen. Logic without contraction as based on inclusion and unrestricted abstraction. *Studia Logica*, 64(3):365–403, 2000. doi:10.1023/A:1005293713265.

Graham Priest. The logic of paradox. *Journal of Philosophical Logic*, 8(1): 219–241, 1979. doi:10.1007/BF00258428.

Graham Priest. *An Introduction to Non-Classical Logic: From If to Is*. Cambridge University Press, 2008.

Willard Van Quine. *Philosophy of Logic*. Prentice-Hall, 2nd edition, 1986.

Greg Restall. A note on naive set theory in *LP*. *Notre Dame Journal of Formal Logic*, 33(3):422–432, 1992. doi:10.1305/ndjfl/1093634406.

David Ripley. Paradoxes and failures of cut. *Australasian Journal of Philosophy*, 91(1):139–164, 2013. doi:10.1080/00048402.2011.630010.

David Ripley. Naive set theory and nontransitive logic. *Review of Symbolic Logic*, 8(3):553–571, 2015. doi:10.1017/S1755020314000501.

Kevin Scharp. *Replacing Truth*. Oxford University Press, 2013.

Peter Schroeder-Heister. Rules of definitional reflection. In *Proceedings of the 8th Annual IEEE Symposium on Logic in Computer Science*, pages 222–232, 1993.

Lionel Shapiro. Deflating logical consequence. *Philosophical Quarterly*, 61 (243):320–342, 2011. doi:10.1111/j.1467-9213.2010.678.x.

Raymond Smullyan. *First-Order Logic*. Dover Books, 1995.

Shawn Standefer. Solovay-type theorems for circular definitions. *Review of Symbolic Logic*, 8(3):467–487, 2015. doi:10.1017/S1755020314000458.

Kazushige Terui. Light affine set theory: A naive set theory of polynomial time. *Studia Logica*, 77(1):9–40, 2004. doi:10.1023/B:STUD.0000034183.33333.6f.

Albert Visser. Semantics and the liar paradox. In D. Gabbay and F. Guenther, editors, *Handbook of Philosophical Logic*, volume 11, pages 149–240. Springer, 2nd edition, 2004.

Zach Weber. Extensionality and restriction in naive set theory. *Studia Logica*, 94(1):87–104, 2010a. doi:10.1007/s11225-010-9225-y.

Zach Weber. Transfinite numbers in paraconsistent set theory. *Review of Symbolic Logic*, 3(1):71–92, 2010b. doi:10.1017/S1755020309990281.

Zach Weber. Transfinite cardinals in paraconsistent set theory. *Review of Symbolic Logic*, 5(2):269–293, 2012. doi:10.1017/S1755020312000019.

P. D. Welch. Games for truth. *Bulletin of Symbolic Logic*, 15(4):410–427, 2009. doi:10.2178/bsl/1255526080.

Stephen Yablo. Definitions, consistent and inconsistent. *Philosophical Studies*, 72(2):147–175, 1993. ISSN 1573-0883. doi:10.1007/BF00989672.

Elia Zardini. Truth without contra(di)ction. *The Review of Symbolic Logic*, 4(4):498–535, 2011. doi:10.1017/S1755020311000177.

Elia Zardini. Naive truth and naive logical properties. *Review of Symbolic Logic*, 7(2):351–384, 2014. doi:10.1017/S1755020314000045.