

Virtuous Circles

G. Aldo Antonelli
Department of Philosophy
University of California – Irvine
IRVINE, CA 92697–4555
aldo@uci.edu

September 1998

For a special issue of the *Journal of the Indian Council of Philosophical Research*
on “Circularity, Definition, and Truth” edited by A. Chapuis and A. Gupta.

1 Some History

Circularity has been the subject of centuries of philosophical bad press. There is a long tradition, going back at least to Aristotle, according to which circularity in all its many forms is characterized as an obstacle to real conceptual and philosophical progress.

In the *Posterior Analytics*, Aristotle takes up the position of those who hold that all knowledge is demonstrable, and, hence, scientific. Such people are said to base their arguments on the fact that some demonstrations are *circular* or *reciprocal* (72b25¹). As Aristotle makes clear in the text, a circular demonstration consists of an argument (form) in which the conclusion is equivalent to one of the premises. But as Aristotle hastens to point out, demonstrations cannot be circular, for the essence of demonstration is to proceed from what is prior to what is posterior, and the same things cannot be both prior and posterior. A circular demonstration has the form ‘if *A* is, then *B* must be;’ and ‘if *B* is, then *A* must be’: “consequently, the upholders of circular demonstration are in the position of saying that if *A* is, *A* must be—a simple way of proving anything” (73a5).

While it is hard to disagree with Aristotle’s first conclusion (that circular demonstrations fail to increase our confidence in their conclusions), there also seems to be a peculiar confusion at work here, a confusion that is still at work in Thomas Aquinas’ influential commentary on the *Posterior Analytics* (see [8]). After summing up and explaining Aristotle’s argument from the *Posterior Analytics*, and coming to the conclusion that a circular argument is (equivalent to one) of the form “if it is *A*, it is *A*,” Aquinas concludes that in this way “it will be easy to demonstrate all things” [8, p. 29].

What is missing, in both Aristotle and Aquinas, is a clear-cut distinction between an argument and its conclusion, a distinction that has forcefully entered modern logic only recently, and most notably with the notion of *conditional proof*. The argument form “if it is *A*, it is *A*” (and any of its variants), while circular, is clearly valid. The fact that the inference “*A*; therefore *A*” can

¹All references are from [15].

be validly carried out for any A does not imply that any proposition A whatsoever can be validly demonstrated. Failure to recognize this fact has led Aristotle (and, with him, Aquinas) to rule out circular demonstrations.

Having noted this peculiarity, we now pass to the second, more interesting attack mounted by Aristotle on the notion of circularity. This attack targets the notion of *circular definitions*. Perhaps the *locus classicus* for this is book I of the *Physics*, where Aristotle is concerned with drawing a distinction between ‘separable’ and ‘non-separable’ attributes. An attribute is separable (roughly) if it is accidental: if it may or may not belong to the subject. For instance, ‘sitting’ is a separable attribute, whereas ‘snub’ is not. The point is that ‘snubness’ contains the definition of ‘nose,’ to which ‘snubness’ is attributed (186b20). Aristotle goes on noticing that if the definition of the whole is contained in the definitions of the attributes making up the definitory formula (the way ‘biped’ or — we add — ‘rational’ is supposed to be an exclusive attribute of ‘man’), then the whole must be a separable attribute, otherwise the definition of ‘man’ would occur in the definition of ‘biped’ (or ‘rational’) — “which is impossible, as the converse is the case” (186b30).

Here Aristotle very clearly condemns definitions that are *circular* in that the *definiendum* in turn occurs in the definitions of the terms making up the *definiens*². Again, it is instructive to look at Thomas Aquinas’ analysis of the passage (in his *Commentary on Aristotle’s “Physics”* [9]). He clearly identifies two *assumptions* Aristotle makes in the *Physics*: the first is the distinction between separable and inseparable attributes. Of these two terms, the latter is “the accident in whose definition is placed the subject in which it is,” as in the case of ‘snub’: “nose is placed in the definition of snub. For the snub is a curved nose” ([9, p. 25]). More importantly from our point of view, the second assumption is that

if certain things are placed in the definition of that which is defined, or in the definition of the things on which the definition depends, then it is impossible that the whole definition of that which is defined be placed in the definition of these certain things ([9, p. 25]).

For if that were to happen, then we would have a circular definition. Why are circular definitions bad? It is interesting to observe that Aquinas provides an *epistemological* ground: for then “one and the same thing would be both prior and posterior, better known and less known” ([9, p. 26]).

Notice that this is the same kind of reason as those provided in order to ban circular reasoning. More importantly perhaps, Aquinas explicitly recognizes that the ban on circular definitions is an *assumption*, and that the Philosopher does not provide any explicit argument for it. And as with all assumptions, this leaves the door open for us to explore what happens when the assumption is dropped.

These pronouncements of Aristotle and Aquinas were extremely influential throughout the history of philosophy. We will not document here the many instances in which they were taken up again (mostly in the form of somebody’s having to defend oneself from the infamous charge of circularity). But we will take notice that, at the other end of the parabola, the same worry about circularity is voiced by Russell and appears to be behind his *vicious-circle principle*:

“Whatever involves *all* of a collection must not be one of the collection”, or, conversely: “If, provided a certain collection had a total, it would have members only definable in terms of that total, then the said collection has no total”. ([19, p. 155]).

As Russell himself readily acknowledges, this principle is purely negative: it does not tell us what collections are to be admitted, but only which ones cannot be allowed. The vicious circle principle

²See [5] for an overview of the terminology and the issues occurring in the theory of definitions.

reflects a preoccupation with *impredicative* definitions, i.e., definitions that involve reference to some totality of which the object being defined is part.

The fairly simple basic intuition underlying the vicious circle principle is then put to work by Russell to give rise to a complex apparatus, the Ramified Theory of Types. The details of the theory need not concern us here, except to remark that the theory is meant to rule out not only circular *definitions*, as was the thrust of the above quote, but to rule out circular *predication* altogether. The Theory of Types regiment the language in such a way as to rule out predication loops. As a consequence, also circular definitions are ruled out.³

The purpose of this paper is two-fold. Our first aim is to show that in spite of these and many other pronouncements, circularity and especially circular definitions are ubiquitous. Second, we will provide at least preliminary evidence that circularity is, in at least a few instances, a desirable phenomenon. Indeed, many fields of human inquiry have to deal with one form or another of circularly characterized phenomena, so that some account of circularity is needed if these are to be accounted for at all. We provide three examples of this: the first is the reconstruction of a classic theorem of set theory, the Cantor-Schröder-Bernstein theorem, that brings the underlying circularity in the foreground; the second is a “revision-theoretic” analysis of general (i.e., total) recursive functions over the natural numbers; and the third is an account of defeasible reasoning that highlights the connections to the theory of truth.

2 Circularity and Fixed Point Equations

As mentioned, in this paper we are going to focus on circular *definitions*, aiming to show that they come in in many fields of human enquiry. In particular, we will take into special consideration circular definitions in the form of explicit equations such as

$$S = \Phi(S), \tag{1}$$

where S can be a set, a relation, or a function, $\Phi(S)$ is an expression in which S itself occurs, and ‘=’ is to be interpreted broadly and appropriately as an identity or an equivalence. Equalities such as the above are banned in the standard modern theory of definition, precisely because of their circularity.⁴

The modern theory of definition imposes two requirements of any expression that is put forward as a definition (in practice, such expressions are often explicit equalities or equivalences, but the requirements are completely general). The first is the requirement of *eliminability*: the theory of definition requires that the *definiendum* be eliminable in every context in the sense that for any expression containing the *definiendum* there must be an equivalent expression not containing it. The second requirement is *conservativeness*: adopting the definition must not allow us to draw any conclusions *not involving the definiendum* that were not inferable before the definition was adopted. Consider for instance a classic definition such as

$$\textit{man} = \textit{rational animal}.$$

³On the set-theoretic side, a theory that takes the intermediate route of ruling out circular definitions but not circular predication (i.e., membership) is Quine’s “New Foundations” — see [6] for more information and further references.

⁴The modern theory of definition originated with the Polish logician S. Lesniewski — again, see [5] for an introductory treatment and further references.

The definition meets the two requirements: the *definiendum* is eliminable, because any expression containing ‘man’ is equivalent (given the definition) to the result of replacing ‘man’ by ‘rational animal’ throughout. The definition is also conservative, since no new conclusions concerning, say, prime numbers or elephants become inferable once the definition is adopted ([5]).

Circular definitions, however, behave quite differently. The most notable feature is that the *definienda* are not eliminable. Even if we replace the left-hand side by the right-hand side of (1) in any given expression A , the resulting expression B still contains occurrences of the *definiendum* S . This is essentially the reason why circular definitions have been ostracized in the modern theory.⁵

But as we mentioned, circularity is ubiquitous. Quite often we inquire about the existence of entities satisfying certain conditions, and these conditions can most naturally be expressed in the form of explicit equations or equivalences such as (1).

A remarkable amount of work has been devoted to the study of such (so-called) “fixed point” equations and the conditions under which they have solutions. The remainder of this paper is devoted to providing examples of this phenomenon, but not before outlining the most important procedures developed for making sense of such circular definitions.

Two main devices have been put to use in order to make sense of circular definitions such as (1). The first device exploits an inductive construction, and although of more limited applicability than the other one, yields solutions that are in many ways ideal. The second device exploits the method of *revision rules* introduced by Gupta, Belnap and Herzberger in connection with the theory of truth and later developed into a full-fledged general theory of circular definitions in Gupta & Belnap [11]. Let us begin with the first one.

Consider again an equation such as (1). Such an equation, as it stands, is a mere formal object, and as such is in need of being interpreted over some given domain. In particular, what we need in order to interpret it is a space of potential extensions of the symbol S . This must be given, or else we have nowhere to go. Once such a domain \mathcal{D} is given, then it is natural to interpret the equation (1) as an operator mapping hypothetical extensions of S into hypothetical extensions of S (i.e. a function from \mathcal{D} into \mathcal{D}). Once a hypothetical extension of S is given as input, we use it to interpret occurrences of S in $\Phi(S)$, evaluate the resulting expression to obtain a hypothetical extension of S as output. And of course, the process can be indefinitely iterated.

Now, depending on the particular features of the *definiens* $\Phi(S)$, it is possible that the hypothetical extensions successively obtained in this way get increasingly “better” (in a sense to be made precise). If this is the case, then the iteration process culminates in a *fixed point*, i.e., a hypothetical extension for S that cannot be further improved. In other words (and with a slight abuse of notation), a fixed point is a hypothetical extension S^* for which, indeed, $S^* = \Phi(S^*)$ holds. The mathematics behind this approach is reviewed in the next section.

The second device that allows us to make sense of circular definitions — Revision Rules — is really a generalization of the inductive construction just mentioned.⁶ In some cases, there is no guarantee that the iteration of the operator $\Phi(S)$ will yield progressively “better” hypothetical

⁵As will be apparent from the treatment of Section 4, one of the claims of the paper is that ordinary recursive definitions are indeed a kind of circular definitions — albeit a kind we know how to treat. In this context, consider the definition of addition by means of the two equations:

$$\begin{aligned}x + 0 &= x, \\x + y' &= (x + y)',\end{aligned}$$

where y' is the successor of y . The definition fails the eliminability criterion, for it does not allow us to eliminate the symbol ‘+’ from, say, $\forall x \forall y (x + y = y + x)$ — although it does allow us to eliminate it from $5 + 7 = 12$.

⁶See Antonelli [3] for a more detailed account of how revision rules can arise naturally by relaxing and generalizing the monotone construction — at least in the case of numerical computation.

extensions of S , and this in turn means that a fixed point might not be eventually reached. This leaves us with the problem of still making sense of the definition.

It is precisely in order to solve this problem in the particular case of truth that Gupta & Belnap developed the Revision Theory of Definition. In this paper we purposely avoid touching upon the topic of truth. Still, we want to take a look at how Revision Rules can be motivated as a general tool for the analysis of circular definitions, including truth.

Part of the reason why circular definitions have been ruled out in the classical theory is the idea that a definition, among other things, should allow us to fix the *extension* of a concept: the preferred way to determine whether an object x falls in the extension of a concept S is to ascertain whether x satisfies the definition of S . If the definiens itself contains occurrences of S , then it might be impossible to determine whether anything satisfies without knowing the extension of S , but that is precisely what we set out to do.

To make things more complicated, sometimes no non-circular definitions are available for a given concept. This is the case of the concept of truth, construed as a predicate of sentences of a given language. If the language in question itself contains the truth predicate then circularity can arise. Tarski [21] suggested that for any sentence φ , the following equivalence (known as a “T-biconditional”) might be regarded as a *partial* definition of truth:

$$t \text{ is true if and only if } \varphi,$$

where t is a term denoting φ (for instance, t could be obtained by enclosing φ in quotes). If φ itself contains occurrences of “is true” the T-biconditional for φ exhibits a form of circularity. As is well known, this circularity can lead to paradoxes, most notably the so-called “liar paradox” arising from the consideration of the T-biconditional for the following sentence λ : λ is not true. This circularity appears to be unavoidable as long as we insist that the language contain its own truth predicate (as is the case for English and other natural languages).

The *Revision Theory of Definition* of Gupta & Belnap does away with the idea that the primary goal of a definition is to fix the extension of the definiendum. Rather, they interpret a circular definition $S = \Phi(S)$ as providing a *rule of revision*, i.e., a way to improve upon hypothetical extensions of S . In this, the revision-theoretic approach is quite similar to the approach behind the inductive construction, but, as already mentioned, it is more general.

The inductive construction exploits the fact that the potential extensions of S get increasingly “better” in order eventually to reach a fixed-point, i.e., a solution for the equation. When the sequence of hypothetical extensions does not exhibit such a behavior, the inductive construction has nothing to say. On the other hand, on the revision-theoretic approach, it’s the *pattern of variation* of the sequence that conveys the information. Should the sequence culminate with a fixed point, all the better, but this is not required in order to provide an analysis of the proposed circular definition.

In other words, what matters is the *revision process* itself. Starting with an arbitrary bootstrapper or initial guess E_0 as to the extension of S , we keep applying the “revision jump” infinitely often: this means that we use the bootstrapper E_0 to interpret occurrences of the definiendum S in the definiens $\Phi(S)$; this gives us a revised extension E_1 , which can in turn be used to interpret occurrences of S in $\Phi(S)$, obtaining E_2 , etc. This poses the problem of what to do at the limit stage. Since there is no guarantee that the hypothetical extensions are increasingly each one better than the previous, at limit stages we cannot simply “collect the accumulated wisdom,” for such a wisdom might convey incompatible answers. Instead we renew our guess, subject to the constraint that the new guess must *cohere* with the previous sequence.

Say that an extension E' *coheres* with the sequence if it contains all items that are eventually in every member of the sequence, and it omits all items that are eventually omitted by every member of the sequence. In other words, for each item x we check if, beginning at some stage or other, x is in every member of the sequence later than stage; in which case we require x to be in E' . Similarly, we check if, beginning at some stage or other, x is outside of every member of the sequence later than that stage; in which case we require x to be outside of E' . If E' contains all x 's eventually contained and omits all x 's eventually omitted, we say that E' coheres with the sequence.

One of the central ideas behind the revision approach is to allow any guess at the limit stage, as long as it coheres with the previous sequence, and then start applying the revision jump again until a next limit stage is reached, a new coherent guess produced, and so on.

Let us emphasize again that what matters here is not whether the process eventually yields a fixed point (which might indeed be the case, as we will see below), but the pattern of variation exhibited for all possible bootstrappers and for all possible coherent guesses at limits. For instance, it is possible that, even if the revision process never converges to a fixed point, there are some objects x from \mathcal{D} that are always eventually put in the extension, independently of the bootstrapper and the different guesses at limit stages. If this is indeed the case, we have at least a partial answer as to the extension of S : we know that it must contain x . Other patterns of behavior are possible and equally interesting. For instance, there might be x 's that are always eventually omitted by the various hypothetical extensions, in which case we also have a partial answer. Or there might be items that stabilize one way for certain bootstrappers and the other way for the remaining ones, exhibiting therefore a milder sort of pathologicity than items that never settle or only settle for certain bootstrappers but not all.

These details are awaiting exploration, and Gupta & Belnap [11] is the obvious starting point. What we want to emphasize here is that there is a rich landscape that was hidden by the inductive construction and that deserves to be mapped and explored. Here we will be content to take a cursory look at three examples of how one might go about making sense of circular definitions, hoping that these examples will be representative of the ubiquity, usefulness, and elegance of such a device.

3 The Cantor-Schröder-Bernstein Theorem

In this section we show that a central theorem in classical set theory can be recast as a fixed point argument. Although this is not new, it is little-known enough still to be interesting. The Cantor-Schröder-Bernstein theorem can be proved in many ways, but the proofs one finds in standard introductory texts in set theory are strikingly cryptic: indeed, the fixed-point argument appears to be by far the most perspicuous proof we know. It will also help make our point that circularity is indeed ubiquitous even in those fields that are supposed to have banned it a long time ago.

It will be useful, before we start, to review some basic facts concerning monotone operators and their fixed-points. Let D be a set and Γ an operator on the power set $\mathcal{P}(D)$ of D . This means that if $X \subseteq D$ then also $\Gamma(X) \subseteq D$. We say that Γ is *monotone* if, given any two subsets X and Y of D , if $X \subseteq Y$ then $\Gamma(X) \subseteq \Gamma(Y)$. A subset Z of D is a *fixed point* of Γ if and only if $\Gamma(Z) = Z$.

The remarkable thing about monotone operators is that they have many fixed points, and among these there is a unique least fixed point. To see why, let us say that a subset X of D is Γ -*closed* if $\Gamma(X) \subseteq X$. Obviously, there is at least one Γ -closed subset of D , viz., D itself. Now let

$$I = \bigcap \{X \subseteq D : X \text{ is } \Gamma\text{-closed}\}.$$

Then I is a fixed point of Γ : indeed, it is the least fixed point of Γ . To establish this, we need to go through a few intermediate steps.

3.1 LEMMA $\Gamma(I) \subseteq \bigcap\{\Gamma(X) : X \text{ is } \Gamma\text{-closed}\}$.

Proof. Pick an arbitrary Γ -closed set X . Then $I \subseteq X$ by definition of I , whence by monotonicity $\Gamma(I) \subseteq \Gamma(X)$. ■

3.2 LEMMA $\bigcap\{\Gamma(X) : X \text{ is } \Gamma\text{-closed}\}$ is a subset of $\bigcap\{X \subseteq D : X \text{ is } \Gamma\text{-closed}\}$.

Proof. To prove the inclusion we need to show that for every Γ -closed Y ,

$$\bigcap\{\Gamma(X) : X \text{ is } \Gamma\text{-closed}\} \subseteq Y.$$

In turn, let a be arbitrary such that $a \in \bigcap\{\Gamma(X) : X \text{ is } \Gamma\text{-closed}\}$; then a belongs to $\Gamma(X)$ for every Γ -closed X and hence in particular $a \in \Gamma(Y)$. But Y is Γ -closed, i.e., $\Gamma(Y) \subseteq Y$, whence $a \in Y$. ■

3.3 LEMMA I is Γ -closed, i.e., $\Gamma(I) \subseteq I$.

Proof. This follows from lemmas 3.1 and 3.2 and the fact that $I = \bigcap\{X : X \text{ is } \Gamma\text{-closed}\}$. ■

3.4 LEMMA I is a fixed point of Γ , indeed a least fixed point.

Proof. Given lemma 3.3, suffices to show $I \subseteq \Gamma(I)$. From $\Gamma(I) \subseteq I$ (by lemma 3.3), we have $\Gamma(\Gamma(I)) \subseteq \Gamma(I)$ by monotony. This shows that $\Gamma(I)$ is Γ -closed, so that $\Gamma(I) \subseteq I$ by definition.

This establishes the existence of at least a fixed point. To see that I is a least fixed point, suppose H is a fixed point of Γ . Then, in particular, $\Gamma(H) \subseteq H$, and hence H belongs to the set $\{X : \Gamma(X) \subseteq X\}$; since I is the intersection of such a set, $I \subseteq H$, as desired. ■

In the particular case in which the monotone operator Γ has a particular form of “continuity,” it is possible to obtain a more concrete grasp of the least fixed point. Let us say that an operator Γ is *continuous* if, for every subset X of D and any $a \in \Gamma(X)$ there is a *finite* subset Y of X such that already $a \in \Gamma(Y)$. (Obviously then, monotony implies that $a \in \Gamma(Z)$ for every $Z \supseteq Y$.) Define a sequence I_0, I_1, \dots of subsets of D by induction on n : first we set $I_0 = \emptyset$ and then (assuming I_n already defined), put $I_{n+1} = \Gamma(I_n)$.

3.5 LEMMA For every $m, n \geq 0$, if $m \leq n$ then $I_m \subseteq I_n$.

Proof. By complete induction on $k = n - m$. If $k = 0$ the result follows immediately. Consider the case for $k > 0$: we need to show that $I_m \subseteq I_n = I_{m+k}$. By induction hypothesis, $I_m \subseteq I_{m+k-1}$. Also by inductive hypothesis, $I_{m+k-1} \subseteq I_{m+k}$, so that $I_m \subseteq I_{m+k} = I_n$, as required. ■

3.6 THEOREM Suppose Γ is monotone and continuous, and let I be its least fixed point. Then $I = \bigcup_{n \geq 0} I_n$.

Proof. First we show $I_n \subseteq I$ by induction on n . The case for $n = 0$ is trivial since $I_0 = \emptyset \subseteq I$. The inductive hypothesis gives $I_n \subseteq I$, whence by monotony $\Gamma(I_n) \subseteq \Gamma(I)$. But $\Gamma(I) = I$ and $\Gamma(I_n) = I_{n+1}$, so $I_{n+1} \subseteq I$ as desired. It follows that $\bigcup_{n \geq 0} I_n \subseteq I$. In this part of the argument we did not make use of the hypothesis of continuity.

Now for the converse inclusion, which does require the hypothesis of continuity. We want to show $I \subseteq \bigcup_{n \geq 0} I_n$. Since I is the intersection of all Γ -closed sets, it suffices to show that $\bigcup_{n \geq 0} I_n$ itself is Γ -closed, i.e.:

$$\Gamma\left(\bigcup_{n \geq 0} I_n\right) \subseteq \bigcup_{n \geq 0} I_n.$$

To establish the last inclusion, let $a \in \Gamma(\bigcup_{n \geq 0} I_n)$. Since Γ is continuous, there is a finite set $A = \{a_1, \dots, a_k\}$ such that $a \in \Gamma(A)$ and $a_i \in \bigcup_{n \geq 0} I_n$ (for $i = 1, \dots, k$). Since $a_1, \dots, a_k \in \bigcup_{n \geq 0} I_n$, there must be integers n_1, \dots, n_k such that each $a_i \in I_{n_i}$ (for $i = 1, \dots, k$). Now let $n^* = \max(n_1, \dots, n_k)$. Then by lemma 3.5, $I_{n_1} \subseteq I_{n^*}, \dots, I_{n_k} \subseteq I_{n^*}$. So $A \subseteq I_{n^*}$; by monotony $\Gamma(A) \subseteq \Gamma(I_{n^*}) = I_{n^*+1} \subseteq \bigcup_{n \geq 0} I_n$. Since $a \in \Gamma(A)$, also $a \in \bigcup_{n \geq 0} I_n$. The generality of a shows that $\Gamma(\bigcup_{n \geq 0} I_n) \subseteq \bigcup_{n \geq 0} I_n$, as required. ■

Having reviewed some basic facts about monotone operators and their fixed points, we now return to the Cantor-Schröder-Bernstein theorem. As mentioned, such a theorem is a main staple of current set-theory, and yet a theorem that is usually proved in ways that are far from transparent. The machinery of fixed points and, more in general, of finding solutions to fixed point equations allows us to give a remarkably swift argument.

The theorem is motivated by cardinality considerations. Even before we formally develop the concept of a cardinal number, we can make sense of cardinality comparisons between sets. For instance we can say that the cardinality of a set A is not greater than (less than or equal to) the cardinality of a set B if there exists an injective function f taking arguments in A and values in B (a function f is *injective* if no two distinct arguments take the same value: if $f(x) = f(y)$ then $x = y$).

Similarly, we can say that A and B have the same cardinality if there is an injective function h from A onto B (a function h is *onto* a set B if for every $b \in B$ there is an argument a such that $h(a) = b$). An injective function from A onto B is also called a *bijection* from A to B .

Of course, we would like that if the cardinality of A is less than or equal to the cardinality of B and conversely the cardinality of B is less than or equal to the cardinality of A , then A and B must have the same cardinality. In view of the proposed analysis of these cardinality notions, it is not obvious how to obtain, given injections $f : A \rightarrow B$ and $g : B \rightarrow A$, a bijection $h : A \leftrightarrow B$. This is exactly the import of the the Cantor-Schröder-Bernstein Theorem. (A word on notation: if $f : A \rightarrow B$ and $C \subseteq A$ then we write $f[C]$ for $\{f(x) : x \in C\}$.)

3.7 THEOREM (CANTOR-SCHRÖDER-BERNSTEIN) Suppose $f : A \rightarrow B$ and $g : B \rightarrow A$ are injections. Then there exists a bijection $h : A \leftrightarrow B$.

Proof. The intuition is to construct the desired bijection h by piecing together the two given bijections, using f on certain elements of A and g on other elements of B in such a way that the choices are consistent.

In other words, we would like to find subsets X and Y of A and B respectively, in such a way that f maps X onto Y and g maps the complement $B - Y$ onto the complement $A - X$, as in Figure 1 (this part of the argument is known as the “Banach decomposition.”) Provided we can

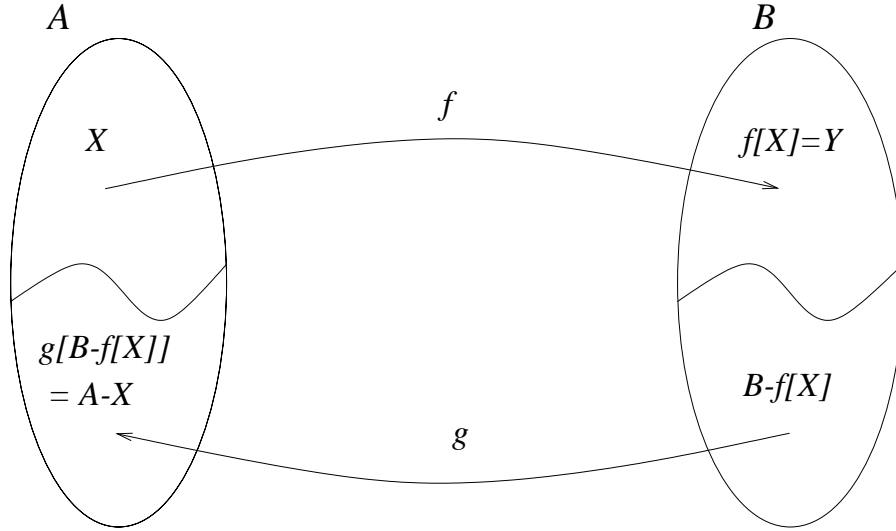


Figure 1: The Banach decomposition.

do that, then h could be explicitly defined thus:

$$h(x) = \begin{cases} f(x) & \text{if } x \in X; \\ z & \text{if } x \notin X \text{ and } g(z) = x. \end{cases}$$

So we need to show that such sets X and Y exist. In fact, since $f[X] = Y$, all we need is a set X such that $g[A - f[X]] = A - X$. In other words we are looking for a solution for the fixed point equation

$$A - g[B - f[X]] = X.$$

Such a solution can be found using the machinery of monotone operators. For clarity, explicitly define an operator Γ taking subsets of A into subsets of A such that for $Z \subseteq A$:

$$\Gamma(Z) = A - g[B - f[Z]].$$

All that is left to do is observe that Γ is monotone:

$$\begin{aligned} Y \subseteq Z &\Rightarrow f[Y] \subseteq f[Z] \\ &\Rightarrow (B - f[Z]) \subseteq (B - f[Y]) \\ &\Rightarrow g[B - f[Z]] \subseteq g[B - f[Y]] \\ &\Rightarrow A - g[B - f[Y]] \subseteq A - g[B - f[Z]] \\ &\Rightarrow \Gamma(Y) \subseteq \Gamma(Z). \end{aligned}$$

If we now let X be a fixed point of Γ , it is clear that X satisfies $A - g[B - f[X]] = X$, as desired. ■

4 General Recursive Functions via Revision Rules⁷

In this section we show how a well-known class of numerical functions, the class of the general (i.e., total) recursive functions can be analyzed in terms of revision rules. The argument takes

⁷A slightly different form of the argument in this section can be found in Antonelli [1]

two steps. First we review Kleene’s [13] analysis of recursive functions as defined by *systems of fixed point equations*, specializing his treatment, however, only to those recursive functions that are everywhere defined (total). Second, we show how to obtain solutions for *all* such systems of equations *simultaneously* by means of revision rules.

The construction given here differs from Kleene not only in its being more specialized (general recursive vs. partial recursive) and in its employing a different construction (revision rules vs. monotone operators). Perhaps the main difference lies in its being a *direct* approach to the total recursive functions. We explain.

As mentioned, the class of the partial recursive functions can be analyzed in terms of fixed-points of monotone operators (see Antonelli [3] for more detailed treatment and further references). It follows that also the total recursive functions are amenable to such a treatment, albeit only via a detour through the partial functions. Such a detour might appear somewhat roundabout if what we are interested in are only total functions. This section shows how to give an analysis of total recursive functions directly, without such a detour, employing ideas and methods from revision theory.

The idea is to proceed as it were “wholesale,” viewing an *entire class* of functionals as specifying a multiple rule of revision that allows us to represent (possibly infinitely) many functions simultaneously. But first we need some auxiliary notions and facts.

4.1 DEFINITION A function $F(\vec{x}, y)$ from \mathbb{N}^k into \mathbb{N} is *regular* if and only if for all \vec{x} there is a y such that (i) $F(\vec{x}, z)$ is defined for all $x \leq y$, and (ii) $F(\vec{x}, y) = 0$.

4.2 DEFINITION The class of *general recursive functions* is defined as the smallest class of functions containing the successor function \mathbf{s} , the constant function $= 0$, and the projection functions $\mathbf{p}_n^i(x_1, \dots, x_n) = x_i$ (for all i, n such that $i \leq n$), and closed under the following operations:

1. *Composition*: if H, G_1, \dots, G_m are general recursive (with arity m and n , respectively), then so is

$$F(x_1, \dots, x_n) = H(G_1(x_1, \dots, x_n), \dots, G_m(x_1, \dots, x_n)).$$

2. *Primitive recursion*: if G and H are general recursive (with arity n and $n + 2$ respectively), then so is

$$F(y, x_1, \dots, x_n) = \begin{cases} G(x_1, \dots, x_n), & \text{if } y = 0; \\ H(y, x_1, \dots, x_n, F(z, x_1, \dots, x_n)), & \text{if } y = \mathbf{s}(z). \end{cases}$$

3. *Least search* applied to regular functions: if G is a *regular* general recursive function of arity $n + 1$, then so is

$$F(x_1, \dots, x_n) = \mu y[G(y; x_1, \dots, x_n) = 0],$$

where, in general, $\mu y[P(y)]$ denotes the least y such that $P(y)$. So $F(x_1, \dots, x_n)$ returns the least y such that $G(y; x_1, \dots, x_n) = 0$ (such a y is guaranteed to exist if G is regular).

The general recursive functions were first developed by Kleene [13], and their special interest lies in the fact that they are *total*, i.e., defined for every argument. Of course, a consequence of this is that other interesting properties will fail, most notably the Enumeration Theorem⁸ (as an

⁸The enumeration theorem for a countable class \mathcal{C} of functions states that there is an assignment of indices from \mathbb{N} to functions in the class (so that \mathcal{C} can be enumerated as, say, F_0, F_1, \dots) and a function F , itself in \mathcal{C} — this is the crucial part — such that for all x : $F(n, x) = F_n(x)$ (and similarly for higher arities).

easy diagonalization argument will reveal). The theorem holds for the *partial recursive* functions (in which there are no constraints on least search), and this might help explain why the latter class of functions is usually preferred to the general recursive as an explication of the intuitive idea of “effectively computable” function. The connection between the two classes is made explicit by the following well-known result, whose proof can be found in any introductory text in recursion theory, e.g., [20, 18].

4.3 THEOREM The class of general recursive functions comprises all and only the partial recursive functions that are total.

We now show that all general recursive functions can be uniformly represented by a unique revision process. This will require that many different revisions can be carried out simultaneously. The idea is as follows. Since the class of general recursive functions is inductively generated, we choose some particular enumeration⁹ of these functions and their definitions:

$$\begin{aligned} F_1(\vec{x}) &= \Phi_1 \\ F_2(\vec{x}) &= \Phi_2 \\ &\vdots \\ F_n(\vec{x}) &= \Phi_n \\ &\vdots \end{aligned}$$

where Φ_i is any one of the right-hand sides of the definition forms allowed for in Definition 4.2, expressed as the conjunction of one or more equations (for instance two equations are required for primitive recursion). Observe that only finitely many symbols F_j ever occur in Φ_i (for $j \leq i$), and F_i itself might be among these.

Next, we uniformly change each definiens Φ_i in this enumeration to Φ_i^* as follows. Let ϕ_i (for $i \in \mathbb{N}$) be countably many new function variables. Then Φ_i^* is just like Φ_i , except that all occurrences of symbols F_j (including occurrences of F_i itself) in it have been changed to occurrences of ϕ_j . If F_i is obtained by least search from, say, the regular function (symbol) F_k , then it might not be obvious how to do the substitution, for the definition

$$F(x_1, \dots, x_n) = \mu y [G(y; x_1, \dots, x_n) = 0]$$

is not, *prima facie*, an equation or a system of equations. However, there are systems of equations that achieve the desired result. An example, adapted and modified from Kleene [13], can be found in Antonelli [3]. We will not give the details here: all that matters is that we have a system of equations defining a function F^* satisfying the condition:

$$F^*(y; \vec{x}) = \begin{cases} y, & \text{if } G(y; x_1, \dots, x_n) = 0; \\ F^*(y + 1; \vec{x}), & \text{if } G(y; x_1, \dots, x_n) \neq 0. \end{cases}$$

Then obviously the desired function F can be obtained by setting

$$F(\vec{x}) = F^*(0, \vec{x}).$$

⁹This by itself does not contradict what we said about the failure of the enumeration theorem, since the enumerating function is not itself general recursive. The subscripts on the F_i 's are *not* indices in the sense of the Enumeration theorem, but only convenient devices to refer to the general recursive functions.

It is important to observe that proceeding in this way we obtain a sequence of *definientia* Φ_i^* (each either an equation or a system of equations), in which the only symbols are symbols for the primitive functions (successor, projection, constant function $= 0$), and the new function variables ϕ_i .

The idea of the construction is then to choose denumerably many bootstrappers ψ_i , and to carry out the revision process simultaneously on all the F_i , at each successor stage interpreting ϕ_i by the result obtained at the previous stage, and taking inferior limits at limit stages. We adopt the notation $(F_i)_\rho^{\psi_i}$ (for ρ an ordinal) to refer to wholesale revision of F_i through ρ stages under initial hypothesis ψ_i , when the revision is carried out simultaneously for all functionals F_j (for $j \in \mathbb{N}$). In order to be rigorous, this needs to be defined by induction on ρ . When $\rho = 0$, then $(F_i)_\rho^{\psi_i}$ is just the bootstrapper function ψ_i . When ρ is a successor, say $\rho = \sigma + 1$, then $(F_i)_\rho^{\psi_i}$ is the function whose value for each argument is obtained by interpreting each occurrence of variables ϕ_j in Φ_i^* by the function $(F_j)_\sigma^{\psi_j}$. Finally, for ρ a limit ordinal, all that we need is to define $(F_i)_{<\rho}^{\psi_i}$. Intuitively this refers to (partial) function whose values stabilize approaching ρ from below:

$$(F_i)_{<\rho}^{\psi_i}(x) = y \iff (\exists \sigma < \rho)(\forall \tau < \rho)(\text{if } \tau \geq \sigma \text{ then } (F_i)_\tau^{\psi_i}(x) = y).$$

We prove in the next theorem that the revision process outlined above, when carried out through the first limit ordinal ω , succeeds in representing all the general recursive functions. A word on notation: in what follows we use f, g, h, \dots , with or without subscripts, as variables, that refer to the functions eventually represented by the revision according to F, G, H, \dots , with or without subscripts.

4.4 THEOREM Let f be general recursive; then there is an elementary functional F such that for all x and ψ , $(F)_{<\omega}^\psi(x) = f(x)$.

Proof. Having done away with subscripts for the sake of conceptual clarity in the statement of the theorem, we re-introduce them right away in its proof for the sake of notational perspicuity. So, as we did above, let $F_i(x)$ be obtained from the definition of f_i . That is, F_i is now to be interpreted as Φ_i^* , defined above. We proceed by induction on the definition of f_i .

The result is trivial if f_i is one of the initial functions, for then $(F_i)_\rho^{\psi_i} = f_i$ already for $\rho = 1$. Suppose that f_i is obtained by composition, say $f_i(x) = f_j(f_k(x))$. By inductive hypothesis there are functionals F_j, F_k such that for all x , $(F_j)_{<\omega}^{\psi_j}(x) = f_j(x)$ and $(F_k)_{<\omega}^{\psi_k}(x) = f_k(x)$. Furthermore, by hypothesis we know that:

$$F_i(x) = F_j(F_k(x)).$$

Now the inductive hypothesis implies that for any x there is a finite stage n (depending on x) at which

$$(F_k)_n^{\psi_k}(x) = f_k(x),$$

and this value is preserved at all later stages. Moreover, there is a finite stage m (depending on $f_k(x)$) such that

$$(F_j)_n^{\psi_j}(f_k(x)) = f_j(f_k(x)),$$

and this value is likewise preserved at all later stages. If we now let $p = \max(m, n)$ we have:

$$\begin{aligned} (F_i)_p^{\psi_i}(x) &= (F_j)_p^{\psi_j}((F_k)_p^{\psi_k}(x)) \\ &= (F_j)_p^{\psi_j}(f_k(x)) \\ &= f_j(f_k(x)) \\ &= f_i(x). \end{aligned}$$

The desired conclusion follows.

Now suppose that f_i is obtained by primitive recursion from functions f_j and f_k :

$$f_i(x, y) = \begin{cases} f_j(x, y), & \text{if } x = 0; \\ f_k(x, y, f_i(x-1, y)) & \text{if } x > 0. \end{cases}$$

We want to show that for every x , $(F_i)_{<\omega}^{\psi_i}(x) = f_i(x)$. We proceed by induction on x . If $x = 0$ then the induction hypothesis tells us that there is a stage n (depending on x, y) at which $(F_j)_n^{\psi_j}(x, y)$ stabilizes at the right value (it “clicks in,” so to speak), so that

$$(F_i)_n^{\psi_i}(x, y) = (F_j)_n^{\psi_j}(x, y) = f_j(x, y) = f_i(x, y).$$

Now suppose that $x > 0$, and assume the result holds for $x - 1$. Then there is n (depending on $x - 1, y$) such that

$$(F_i)_n^{\psi_i}(x - 1, y) = f_i(x - 1, y),$$

and this is preserved at all later stages. Moreover, there is also m (depending on x, y and $f_i(x - 1, y)$) such that

$$(F_k)_m^{\psi_k}(x, y, f_i(x - 1, y)) = f_k(x, y, f_i(x - 1, y)),$$

and this value is preserved at all later stages. Now let $p = \max(m, n + 1)$. Then we have:

$$\begin{aligned} (F_i)_p^{\psi_i}(x, y) &= (F_k)_p^{\psi_k}(x, y, (F_i)_p^{\psi_i}(x - 1, y)) \\ &= (F_k)_p^{\psi_k}(x, y, f_i(x - 1, y)) \\ &= f_k(x, y, f_i(x - 1, y)) \\ &= f_i(x). \end{aligned}$$

The desired conclusion follows. Now we get to the crucial case. Let f_j be a regular function of two arguments, and suppose that f_i is obtained by least search from it:

$$f_i(x) = \mu y [f_j(x, y) = 0].$$

Now fix x ; since f_j is regular, there is y such that $f_j(x, y) = 0$. By inductive hypothesis, there is a stage n such that

$$(F_j)_n^{\psi_j}(x, y) = f_j(x, y) = 0.$$

Given our definition of Φ_i^* , $(F_i)_n^{\psi_i}$ starts an upward search for the first y such that $(F_j)_n^{\psi_j}(x, y) = 0$. Since f_j is regular by hypothesis, and by inductive hypothesis $(F_j)_n^{\psi_j}(z) = f_j(z)$ for all $z \leq x$, the search succeeds, and the conclusion follows. \blacksquare

5 Defeasible Reasoning¹⁰

Finally, we come to our last example of circularly defined notions: defeasible inference. A number of so-called *non-monotonic logics* were developed over the last twenty years to provide a formal framework within which to model cognitive phenomena such as *defeasible inference* and *defeasible knowledge representation*, i.e., to provide a formal account of the fact that reasoners can reach conclusions tentatively, reserving the right to retract them in the light of further information.

¹⁰The material in this section has an overlap with Antonelli [2]

These logics represent one of the most important recent developments both in logic and in Artificial Intelligence.

In this section we present a particular inferential formalism, i.e., a simplified version of Reiter’s default logic [16]. After providing the fundamental definitions and illustrating the basic facts concerning this version of default logic, we show how to provide a somewhat alternative account of this form of defeasible reasoning, in such a way as to capture the inherent circularity and to bring it in line with the analysis of other circular phenomena presented in this pages. It should be mentioned that the considerations in this section aim to be motivational; a complete treatment of the formal details can be found in Antonelli [4, 7].

Default logic is intended to represent defeasible inferences of the the form: “*If Tweety is a bird, then in the absence of information to the effect that Tweety is a penguin, infer that Tweety flies*”. The inference is defeasible because when we acquire further information to the effect that Tweety is indeed a penguin, the conclusion of the inference is retracted. Such a defeasible inference rule can be represented by a “default” — i.e., a defeasible inference rule — such as:

$$\frac{\text{Tweety is a bird} : \text{Tweety is not a penguin}}{\text{Tweety flies}}.$$

More in general, a default δ is an expression of the form

$$\frac{\zeta : \eta}{\theta},$$

where ζ, η, θ are sentences from a given language \mathcal{L} , to be specified below. The expressions ζ, η, θ are called the *pre-requisite*, the *justification*, and the *conclusion* of δ , respectively. The intuitive meaning of δ is that if ζ is known, and η is consistent with what is known, then we are entitled to infer θ (“by default”). To say that η is consistent with what is known is to say that we have no information to the effect that not- η is true. If this is the case, then the default is “fired” or “triggered,” and its conclusion added to our knowledge base.

The basic problem one encounters when trying to represent defeasible inference by means of defaults is that we want η to *remain consistent* even *after* the default is fired. That is to say, there is an inherent circularity in defeasible reasoning that needs to be accounted for if we want to give a precise formal account. Default logic gives such an account by providing a solution to basic problem, using the notion of *extension*. But before we can give the details we need a few more definitions to fix concepts and terminology.

We are going to work with a classical propositional language \mathcal{L} , of the usual sort. In particular, \mathcal{L} is obtained from propositional constants p, q, r, \dots using the propositional connectives $\&, \vee, \neg, \rightarrow$. We will employ the classical notion of logical consequence, according to which a sentence φ is a logical consequence of a set of sentences S if and only if φ is true on any truth-value assignment on which all sentences in S are true. Similarly, φ is inconsistent with a set S of sentences if and only if $\neg\varphi$ is a consequence of S .

There are several important kinds of defaults. In the particular case in which a default δ has no pre-requisite or, equivalently, its prerequisite is a tautology, we say that δ is *categorical*. Reiter [16] also singles out the notion of a *normal* default: this is a default whose justification is the same as its conclusion. Similarly, Reiter & Criscuolo [17] define *semi-normal* defaults to be defaults of the form

$$\frac{\zeta : \eta \& \theta}{\theta},$$

in which the conclusion occurs as a conjunct in the pre-requisite.

Given the above construal of defaults, we can introduce Reiter’s notion of a default theory. By a *default theory* we understand a pair (W, Δ) , where W is a set of propositional axioms in \mathcal{L} (a “world-description”) and Δ is a *finite* set of defaults $\delta_1, \dots, \delta_n$. A default theory (W, Δ) is *categorical*, *normal*, or *semi-normal*, according as all defaults in Δ are categorical, normal, or semi-normal. As we will see, categorical default theories form a natural and well-behaved class.

Let (W, Δ) be a default theory, S a set of \mathcal{L} -sentences, and δ a default from Δ . Then we say:

1. δ is *admissible* in S if and only if the prerequisite of δ is a consequence of $S \cup W$;
2. δ is *conflicted* in S if and only if the conclusion of δ is inconsistent with $S \cup W$;
3. δ is *pre-empted* in S if and only if the justification of δ is inconsistent with $S \cup W$.

If Γ is a set of defaults, we say that δ is admissible, conflicted, or pre-empted in Γ according as δ is admissible, pre-empted or conflicted in the set of all conclusions of defaults from Γ .

We now introduce Reiter’s notion of an *extension* (here referred to as a *classical extension*¹¹) for a *categorical* default theory (see [7] for the general case). Intuitively, an extension Γ — for a categorical default theory (W, Δ) — is a maximal set of defaults that are triggered relative to the knowledge base obtained collecting the conclusions of defaults in Γ . Notice how the definition refers back to itself in a circular fashion.

More formally, we say that a set Γ of defaults is a *classical extension* for a categorical default theory (W, Δ) if and only if it satisfies the following equation:

$$\Gamma = \{\delta \in \Delta : \delta \text{ is not pre-empted in } \Gamma\}.$$

It is well known that if Γ is a classical extension for (W, Δ) then the set of conclusions of defaults in Γ is inconsistent with W if and only if W is already inconsistent. Similarly, if Γ is a classical extension for (W, Δ) and δ is neither conflicted nor pre-empted in Γ , then δ is in Γ .

Given the above definition of an extension for a (categorical) default theory, it is worth noting that it is not obvious how to determine whether a default theory has an extension and, if so, how to construct one. First of all, not all default theories have extensions, as we can see by considering a default theory (W, Δ) where W is empty and Δ contains only the default

$$\delta = \frac{:p}{\neg p},$$

where p is atomic. Suppose Γ were an extension for the theory: then there are two cases, according as δ is in Γ or not. If it is, then the justification of δ , i.e., p , is inconsistent with the set of conclusions of defaults from Γ , so that δ is pre-empted in Γ and therefore cannot be in Γ after all (given that Γ satisfies the equation defining extensions). On the other hand, if δ is not in Γ , then there are no defaults in Γ so that no atomic sentence can be inconsistent with the set of conclusions of defaults in Γ , and in particular δ is not pre-empted in Γ ; so δ is in Γ after all. This shows that, under the hypothesis that Γ is an extension, δ is in Γ if and only if it is not in Γ , an obvious contradiction. We conclude that no extensions exist.

On the other hand, default theories can have multiple extensions. To see this, consider the default theory (W, Δ) , where W is empty and Δ comprises the two defaults:

$$\delta_1 = \frac{:p}{\neg q} \quad \text{and} \quad \delta_2 = \frac{:q}{\neg p}$$

¹¹We will later introduce a different kind of extension.

It is immediate to check that no extension Γ can trigger both defaults, otherwise they would both be pre-empted in Γ . Moreover, any extension for the theory must trigger at least one of them: for if neither belonged to the extension, neither would be pre-empted in such an extension and therefore would have to belong to the extension after all. So there are two classical extensions Γ_1 and Γ_2 , each triggering one default and pre-empting the other one.

This is a general phenomenon: default theories can have multiple classical extensions, that are all incomparable with respect to the subset relation. It follows, in particular, that there cannot be a unique minimal (classical) extension of a default theory. This fact turns out to be problematic if we are interested in defining a notion of defeasible consequence for a default theory, in analogy to the classical notion of logical consequence. Indeed, it would seem that we ought to be interested in such a notion, if default logic is to be a *logic*, and not only a representation device for defeasible inference rules. In particular, we would like to define a relation \vdash that a default theory has to the sentences that are warranted by it, interpret such a relation as *defeasible consequence*, and write, say, $(W, \Delta) \vdash \varphi$ just in case φ is a defeasible consequence of (W, Δ) .

In general, given a default theory having multiple classical extensions, there are several ways in which we can define a relation \vdash : (i) we can decide to be *credulous* and say that $(W, \Delta) \vdash \varphi$ precisely when φ follows from (the set of conclusions of defaults in) *some* extension of (W, Δ) ; (ii) we can arbitrarily pick an extension Γ among the many possible, and decide that $(W, \Delta) \vdash \varphi$ just in case φ follows from (conclusions of defaults in) Γ ; or (iii) we can be *skeptical* and say that $(W, \Delta) \vdash \varphi$ just in case φ follows from (conclusions of defaults in) *all* extensions of (W, Δ) .

All three alternatives have drawbacks. Alternative (ii) is not acceptable for the arbitrariness of our choice of Γ . Alternative (i) can lead us sometimes to endorse contradictory statements (as in the case of Γ_1 and Γ_2 above). Alternative (iii) is the one that best resonates with certain intuitions about defeasible reasoning, e.g., the fact that defeasible reasoners should be cautious in drawing their inferences (see Horty *et al.* [12] for a general argument in favor of skepticism in defeasible reasoning), but goes about it the wrong way. It is certainly a funny way to be skeptical to generate *all* possible extensions of a theory and then take the intersection. If feasibility of computation is an issue at all, this is by far the least resource-oriented approach, and certainly not a satisfactory implementation of skepticism.

It is also important to observe that classical extensions for default theories cannot be constructed (when they exist) by means by any *cumulative process*, of the sort in which defaults are successively assessed for some kind of property which can guarantee their belonging to the extension being constructed. On the contrary, we first have to “guess” a set Γ of defaults and then check that it does indeed satisfy the equation defining extensions.

An important property of this notion of classical extension is its intrinsic “two-valued” character: by this we refer to the fact that such an extension contains the consequences of a maximal set of defaults whose justifications are consistent with the extension itself. In other words, the triggering of a default can only be prevented if its justification is explicitly refuted. The approach to default logic that will be presented below seeks to circumvent this restriction by identifying a “three-valued” notion of extension for default logic, analogous to the one put forward in Antonelli [4] for defeasible inheritance networks. In particular, we will provide a notion of extension that (i) subsumes the classical one and (ii) according to which *any* default theory has an extension. As will see, our notion of extension has other pleasant mathematical properties.

Inspiration for the present approach derives from Kripke’s work [14] in the theory of truth. In turn, Kripke’s approach exploits a construction originally due to Kleene in connection with the analysis of partial recursive functions (see Feferman [10] for a strengthening of the Kripke construction as well as a historical appraisal of its predecessors). Kripke considers a language

containing its own truth predicate, avoiding paradox by switching to a three-valued setting. In such a setting, a sentence is true, false or indeterminate according as its negation is respectively false, true or indeterminate. This intuition (extended to the other connectives in the language) allows one to carry out an inductive construction not too dissimilar — as we will see — from the one that gives minimal general extensions for default theories. By weakening the semantics Kripke achieved a better-behaved model of how natural language can contain its own truth predicate, and the maneuver carried out here is similar in inspiration.

Another trait of the present proposal is that it embodies a “cautious” or “skeptical” approach to defeasible reasoning, which in turn is substantiated in our three-valued notion of *general extension*. The main idea is that A default δ can be prevented from being triggered if and only if it is either explicitly conflicted or *potentially pre-empted*. Contrast this to Reiter’s notion of extension, in which δ can be prevented from being triggered if and only if it is *explicitly pre-empted*. This is a crucial difference, within which lies the more cautious nature of general extensions. Moreover, again in contrast to Reiter’s extensions, extensions of the sort proposed here can be obtained as the limit of a genuine inductive or cumulative construction. This feature makes general extensions somewhat better behaved mathematically, e.g., by imposing on general extensions a non-trivial algebraic structure. As a by-product, we have that many default theories have a *least* general extension. Again in contrast, as we remarked, any two of Reiter’s extensions for a given default theory are incomparable with respect to the subset relation.

We are now ready to introduce our notion of “general extension” for (categorical) default theories. A classical extension plays a two-fold role: on the one hand it explicitly comprises the defaults that are triggered (and these are the defaults that belong to the extension), and on the other hand it implicitly specifies the set of defaults that are ruled out (because they do not belong to the extension). The basic intuition at the basis of the notion of a general extension is that these two roles are going to be decoupled, by explicitly identifying a set of defaults as triggered as well as explicitly identifying a set of defaults as ruled out. This leads to the following definition.

A *general extension* for a categorical default theory (W, Δ) is a pair (Γ^+, Γ^-) of sets of defaults from Δ , simultaneously satisfying the following two equations:

$$\begin{aligned}\Gamma^+ &= \{\delta : \delta \text{ is neither conflicted in } \Gamma^+ \text{ nor pre-empted in } \Delta - \Gamma^-\}; \\ \Gamma^- &= \{\delta : \delta \text{ is conflicted or pre-empted in } \Gamma^+\}.\end{aligned}$$

In other words, Γ^+ is the set of all defaults that are neither conflicted in Γ^+ nor pre-empted in $\Delta - \Gamma^-$, while Γ^- is the set of all defaults that are either conflicted or pre-empted in Γ^+ . Notice that before allowing a default in Γ^+ we make sure that it is not potentially pre-empted, i.e., that it is not pre-empted by any other defaults *that have not already been explicitly ruled out*.

It is possible to show that general extensions indeed generalize the notion of classical extension, and that general extensions always exist. Given a categorical default theory (W, Δ) having a classical extension Γ^+ , we can obtain a general extension for (W, Δ) by putting

$$\Gamma^- = \{\delta : \delta \text{ conflicted or pre-empted in } \Gamma^+\}.$$

Then (Γ^+, Γ^-) is a general extension for (W, Δ) . (The reader is referred to [7] for a proof.)

The second important observation that we need to make is that every categorical default theory has a general extension. This is particularly simple to see in the case of categorical, *semi-normal* default theories, but it is also true in the general case. Given a categorical, semi-normal theory (W, Δ) , one can define the operator \mathfrak{G} taking pairs of subsets of Δ into pairs of subsets of Δ :

$$\mathfrak{G}(\Gamma_1, \Gamma_2) = (\Theta_1, \Theta_2),$$

where:

$$\begin{aligned}\Theta_1 &= \{\delta : \delta \text{ is not pre-empted in } \Delta - \Gamma_2\}; \\ \Theta_2 &= \{\delta : \delta \text{ is conflicted or pre-empted in } \Theta_1\}.\end{aligned}$$

(Notice that (Θ_1, Θ_2) depends on Γ_2 only and not also on Γ_1 .) There is an important sense in which \mathfrak{G} is monotone. Define the following ordering \leq on pairs of sets of defaults from Δ : say that $(\Gamma_1, \Gamma_2) \leq (\Pi_1, \Pi_2)$ if and only if $\Gamma_1 \subseteq \Pi_1$ and $\Gamma_2 \subseteq \Pi_2$. Then \mathfrak{G} is monotone with respect to \leq , in the sense that if $(\Gamma_1, \Gamma_2) \leq (\Pi_1, \Pi_2)$ then $\mathfrak{G}(\Gamma_1, \Gamma_2) \leq \mathfrak{G}(\Pi_1, \Pi_2)$. As in section 3, one can show that any such operator has a fixed point, and then proceed to verify that any fixed point is indeed a general extension (the latter part requires the hypothesis that Δ is semi-normal).

Even when Δ is not semi-normal, it is still possible to prove that extensions always exists, but the proof is somewhat more complicated as it involves a *non-deterministic* inductive construction (again, see [7] for the details).

The net effect of the restriction to semi-normal theories is to eliminate any residual non-determinism from the construction. For then no default can be conflicted without being already pre-empted, since the conclusion of the default occurs as a conjunct in the justification: any conflicted default would already be pre-empted.

Let us consider a few examples. Consider first the default theory (W, Δ) , where W is empty and Δ comprises the two defaults:

$$\frac{:p}{\neg q} \quad \text{and} \quad \frac{:q}{\neg p}.$$

As we have seen, this theory has two classical extensions, according to which default is triggered. In addition to these, the theory has one general extension, in which no default is triggered and none is ruled out. This is the unique least extension of the theory.

Similarly, the default theory in which W is empty, and Δ comprises only the default

$$\frac{:p}{\neg p},$$

has one general extension, namely, (\emptyset, \emptyset) , in which the default is neither triggered nor ruled out. We already know that such a theory has no classical extensions.

It might seem that minimal extensions are quite uninteresting, given that in the ones we have seen so far no defaults are triggered and none are ruled out. It turns out that this is not always the case. Consider for instance the theory where W is empty, but Δ comprises the defaults

$$\frac{:p}{\neg p} \quad \text{and} \quad \frac{:q}{r}$$

The second default has nothing to do with the first, and so it should be triggered in any extension. However, the presence of the first default prevents the theory from having any classical extension. In contrast, the theory does have one general extension, namely

$$\left(\frac{:q}{r}, \emptyset \right).$$

Indeed, the first default cannot be triggered in any extension, but there is no obstacle that prevents triggering the second, which is in fact contained in all extensions of the theory, including any minimal one.

Now, consider a different sort of theory, in which W is empty and $\Delta = \{\delta_1, \dots, \delta_n\}$, for some $n > 1$. Suppose also that for all k such that $1 \leq k < n$,

$$\delta_k = \frac{:p_k}{\neg p_{k+1}},$$

whereas

$$\delta_n = \frac{:p_n}{\neg p_1}.$$

So, Δ is a sequence of defaults, each one of which pre-empts the next one, and the last one of which pre-empts the first. It is easy to check that this theory has no classical extensions if n is odd. (It has, of course one general extension even when n is odd, namely (\emptyset, \emptyset) .) On the other hand, if n is even, say $n = 2m$, beside the general extension (\emptyset, \emptyset) , there are two classical extensions, namely

$$(\{\delta_{2k+1} : 0 \leq k < m\}, \{\delta_{2k+2} : 0 \leq k < m\}),$$

and

$$(\{\delta_{2k+2} : 0 \leq k < m\}, \{\delta_{2k+1} : 0 \leq k < m\}).$$

For the details the reader is referred to [7, 4].

Finally, consider the theory (W, Δ) where again W is empty, and Δ contains the two defaults

$$\frac{:q}{p} \quad \text{and} \quad \frac{:q}{\neg p}.$$

This theory has two classical extensions (according to which default is fired), and these are also the only two general extensions. This example is instructive because it shows that even *minimal* general extensions need not be unique. It is however easy to change the above theory in such a way that no default is conflicted in any extension unless it is already pre-empted in it. This amounts to turning the theory into a *semi-normal* default theory by a maneuver first identified by Reiter & Crisculo [17]: by replacing Δ by the set of defaults

$$\frac{:q \ \& \ p}{p} \quad \text{and} \quad \frac{:q \ \& \ \neg p}{\neg p}.$$

This theory gains a unique minimal general extension in which no default is triggered.

We have already remarked that classical extensions do not seem to allow us to define a notion of defeasible consequence in a natural way. The case is quite different for general extensions. As shown in [7], even when minimal extension are not unique, it is still possible to define a well-behaved relation \sim of defeasible consequence.

Things are even better when we consider semi-normal default theories (and, given any theory, we can always switch to a semi-normal theory in the manner advocated by Reiter & Crisculo [17]). Insofar as existence of *classical* extensions is concerned, semi-normal theories do not fare any better than arbitrary default theories, since semi-normal default theories need not have any classical extension (see [17, §3]). But the switch becomes critical from the point of view of general extensions. As we have seen, semi-normal default theories always have a *unique* minimal general extension. This allows us, for any given semi-normal (categorical) default theory (W, Δ) , to define $(W, \Delta) \sim \varphi$ if and only if φ is a logical consequence of conclusions of defaults in Γ , where Γ is the unique minimal extension of (W, Δ) .

What we have been saying so far applies to categorical default theories. However, this restriction is unessential, and can be lifted at the cost of complicating our definitions and proofs

somewhat. The reader is referred to [7] for a full treatment of arbitrary default theories. (The existence of unique minimal extensions for semi-normal categorical default theories extends smoothly to semi-normal default theories.) The point remains that one can account for the inherent circularity characterizing defeasible reasoning by means of the notion of general extension for a default theory, in such a way as to allow: (i) every default theory to have an extension; (ii) minimal extensions to be obtained by means of an inductive (cumulative) process; and (iii) define a well-behaved relation of defeasible consequence for default theories.

6 Epilogue

We started out noticing that Thomas Aquinas correctly recognized that the ban on circular definitions that Aristotle introduced was an *assumption*, motivated on epistemological grounds but not otherwise necessitated. This opened the way for the investigation of what might happen, once such a ban is lifted.

We can then see that circularity is in many ways a desirable trait, which allows us to give a precise and rigorous account of many interesting phenomena, which would otherwise remain outside the reach of formal enquiry. Of course this requires that we have methods and procedures adequate for such an account.

In this paper we aimed to show that not only such methods and procedures are available, but that the phenomena that require them are truly ubiquitous, ranging from classical theorems of set theory, to the definition of computable functions, to the analysis of defeasible reasoning. Of course, we could only supply but a few examples. But further instances are not hard to find, now that we know where to look.

References

- [1] G. Aldo Antonelli, **Revision Rules. An Investigation into Non-Monotonic Inductive Definitions**, doctoral dissertation, University of Pittsburgh, VII–108 pp., Pittsburgh, Penn., 1992.
- [2] G. Aldo Antonelli, *Defeasible Reasoning as a Cognitive Model*, in Krister Segerberg (ed.), **The Parikh Project. Seven papers in honour of Rohit**, Uppsala Prints and Preprints in Philosophy, 1996 n.18, Dept. of Philosophy, Uppsala University.
- [3] G. Aldo Antonelli, *What's in a Function?*, **Synthése** 107 n. 2, 1996, pp. 167–204 (*Math. Rev.* 97k03053).
- [4] G. Aldo Antonelli, *Defeasible Inheritance on Cyclic Networks*, **Artificial Intelligence** v. 92 (1997), pp. 1–23 (*Math. Rev.* 98a68170).
- [5] G. Aldo Antonelli, *Definition*, in Edward Craig (ed.), **The Routledge Encyclopedia of Philosophy**, Routledge, London, 1998.
- [6] G. Aldo Antonelli, *Conceptions and paradoxes of Sets*, to appear in **Philosophia Mathematica**.
- [7] G. Aldo Antonelli, *A Directly Cautious Theory of Defeasible Consequence for Default Logic via the notion of General Extension*, submitted to **Artificial Intelligence**.

- [8] Thomas Aquinas, **Commentary on the Posterior Analytics of Aristotle**, transl. F.R. Larcher, O.P., Magi Books, Albany, NY 1970.
- [9] Thomas Aquinas, **Commentary on Aristotle's "Physics"**, transl. R.J. Blackwell, R.J. Spath, & W.E. Thirlkel, Yale University Press, New Haven, CT 1963.
- [10] Solomon Feferman, *Toward Useful Type-Free Theories, I*, in R. Martin (ed.), **Recent Essays on Truth and the Liar Paradox**, Oxford University Press, Oxford 1984, pp. 237–287.
- [11] Anil Gupta & Nuel D. Belnap, **The Revision Theory of Truth**, MIT Press, Cambridge, Mass. 1993.
- [12] John F. Horty, Richmond H. Thomason, & David S. Touretzky, *A Skeptical Theory of Inheritance in Nonmonotonic Semantic Networks*, **Artificial Intelligence** 42 (1990), pp. 311–48.
- [13] Stephen C. Kleene, **Introduction to Metamathematics**, van Nostrand, Princeton, New Jersey, 1952, X–550 pp.
- [14] Saul Kripke, *Outline of a Theory of Truth*, **The Journal of Philosophy** 72 (1975), pp. 690–716.
- [15] Richard McKeon (ed.), **The Basic Works of Aristotle**, Random House, New York, 1941, XL–1487 pp.
- [16] Raymond Reiter, *A Logic for Default Reasoning*, **Artificial Intelligence** 13 (1980), pp. 81–132.
- [17] Raymond Reiter & Giovanni Criscuolo, *On Interacting Defaults*, **Proceedings of the Seventh International Joint Conference on Artificial Intelligence**, Vancouver, B.C. (1981), pp. 270–276.
- [18] Hartley Rogers, Jr., **Theory of Recursive Functions and Effective Computability**, MIT Press, Cambridge, Mass., 1967, XXI–482 pp.
- [19] Bertrand Russell, *Mathematical Logic as Based on the Theory of Types*, **American Journal of Mathematics** v. 30, 1908 pp. 222–62, reprinted in Jean van Heijenoort (ed.), **From Frege to Gödel**, Harvard University Press, Cambridge, Mass. 1967, pp. 150–82.
- [20] Robert I. Soare, **Recursively enumerable Sets and Degrees**, Springer-Verlag, Berlin and New York, 1987, XVIII–437 pp.
- [21] Alfred Tarski, *The Semantic Conception of Truth*, **Philosophy and Phenomenological Research** 4 (1944), pp. 341–376; reprinted in L. Linsky (ed.), **Semantics and the Philosophy of Language: A Collection of Readings**, University of Illinois Press, Champaign, Ill. 1952.