

## Dangerous Reference Graphs and Semantic Paradoxes

Landon Rabern · Brian Rabern ·  
Matthew Macauley

Received: 29 October 2010 / Accepted: 28 August 2012 / Published online: 2 October 2012  
© Springer Science+Business Media B.V. 2012

**Abstract** The semantic paradoxes are often associated with self-reference or referential circularity. Yablo (*Analysis* 53(4):251–252, 1993), however, has shown that there are infinitary versions of the paradoxes that do not involve this form of circularity. It remains an open question what *relations of reference* between collections of sentences afford the structure necessary for paradoxicality. In this essay, we lay the groundwork for a general investigation into the nature of reference structures that support the semantic paradoxes and the semantic hypodoxes. We develop a functionally complete infinitary propositional language endowed with a denotation assignment and extract the reference structural information in terms of graph-theoretic properties. We introduce the new concepts of *dangerous* and *precarious* reference graphs, which allows us to rigorously define the task: *classify the dangerous and precarious directed graphs purely in terms of their graph-theoretic properties*. Ungroundedness will be shown to fully characterize the precarious reference graphs and fully characterize the dangerous *finite* graphs. We prove that an undirected graph has a dangerous orientation if and only if it contains a cycle,

---

L. Rabern  
School of Mathematical and Statistical Sciences, Arizona State University,  
Tempe, AZ 85287, USA  
e-mail: landon.rabern@gmail.com

B. Rabern (✉)  
School of Philosophy, RISS, The Australian National University,  
Canberra ACT 0200, Australia  
e-mail: brian.rabern@gmail.com

M. Macauley  
Department of Mathematical Sciences, Clemson University, Clemson,  
SC 29634 USA  
e-mail: macaule@clemson.edu

providing some support for the traditional idea that cyclic structure is required for paradoxicality. This leaves the task of classifying danger for infinite acyclic reference graphs. We provide some compactness results, which give further necessary conditions on danger in infinite graphs, which in conjunction with a notion of *self-containment* allows us to prove that dangerous acyclic graphs must have infinitely many vertices with infinite out-degree. But a full characterization of danger remains an open question. In the appendices we relate our results to the results given in Cook (J Symb Log 69(3):767–774, 2004) and Yablo (2006) with respect to more restricted sentences systems, which we call  $\mathcal{F}$ -systems.

**Keywords** Paradox · Hypodox · Reference structure · Circularity · Ungroundedness · Yablo’s paradox · Liar paradox · Graph theory · Dangerous · Precarious ·  $\mathcal{F}$ -system · Kernel

The semantic paradoxes are often associated with self-reference or referential circularity. Yablo [13],<sup>1</sup> however, has shown that there are infinitary versions of the paradoxes that do not involve this form of circularity.<sup>2</sup> The attempts to purge the semantic antinomies by banning self-reference or by constructing sophisticated hierarchies only eliminate the class of paradoxes that rely on the circular reference structures—if cyclical reference is not essential to the semantic paradoxes, then the acyclical paradoxes remain unscathed. It remains an open question what *relations of reference* between collections of sentences afford the structure necessary for paradoxicality. Since “circularity” has traditionally been assumed to be essential, this issue has been underrepresented in the literature on truth and semantic paradoxes<sup>3</sup>—but it is clear that no such theory can lay claim to comprehensiveness until this question is answered. The resolution of this general question, then, has great import for philosophical and mathematical accounts of truth.

In this essay, we lay the groundwork for a general investigation into the nature of reference structures that support the semantic paradoxes (e.g. the Liar and Yablo’s paradox) and the semantic hypodoxes (e.g. the Truth-teller).

<sup>1</sup>An early version of Yablo’s  $\omega$ -paradox can be found in [12, p. 340].

<sup>2</sup>Priest [8] argues that Yablo’s paradox actually does involve a form of circularity. Cook [3] argues that the (original) quantificational version of Yablo’s paradox does indeed involve this form of “circularity” but Cook insists that this kind of circularity (which [3] classifies as “weak fixed point circularity”) is ubiquitous in the language of arithmetic and therefore should not be thought of as the “culprit” involved in the paradox. Cook [3], then, goes on to show how to get rid of this weak form of circularity by moving to an infinitary language—these are constructed by replacing universal quantification with infinite conjunction. The existence of fixed points, in this sense, seems to be an artifact of encoding the paradox in a language that is too weak to support genuine infinitary constructions (e.g. the language of arithmetic)—in this way the only resource available is “potential infinities” in the form of recursive definitions that are circular by their very nature. By *Yablo’s paradox* we mean the infinitary version that does not involve (strong or weak) “fixed point circularity”.

<sup>3</sup>Although some preliminary investigations into paradox supporting structures have been conducted in [2, 11, 13, 14] on a special class of restricted languages (see Appendix D).

To this end, in Section 1 we develop a functionally complete propositional language endowed with reference structure, which provides the *sentence systems* that are susceptible to paradox (hypodox). For a given sentence system we demonstrate how to extract the reference structural information in terms of graph-theoretic properties and introduce the notions of *dangerous* and *precarious* reference graphs (Section 2). This allows us to rigorously define the task: *classify the dangerous and precarious directed graphs purely in terms of their graph-theoretic properties*.

We make some significant progress towards this goal. Some interesting and useful danger (precarity) preserving operations are discussed in Section 3, including subdivision, smoothing, and unwinding. In Section 4 we introduce the notion of “ungroundedness”,<sup>4</sup> which will be shown to fully characterize dangerous *finite* graphs and fully characterize the precarious reference graphs. Since there are acyclic infinite reference configurations, which in spite of their *ungroundedness*, are unable to support paradoxes, it seems that the essential nature of paradox supporting reference patterns is characterized neither in terms of circularity nor ungroundedness. Nevertheless, in Section 6 we also prove that an undirected graph has a dangerous orientation if and only if it contains a cycle. So there remains some sense in which cyclic structure is required for paradoxicality—this result requires further philosophical interpretation.

This leaves us the task of classifying danger for infinite acyclic reference graphs. In Section 5 we provide some compactness results, which give further necessary conditions on danger in infinite graphs. Using the compactness results in conjunction with a notion of *self-containment* we prove that dangerous acyclic graphs must have infinitely many vertices with infinite out-degree. Overall we issue an interesting set of necessary and sufficient conditions on the danger of infinite reference graphs. But a full characterization of danger remains an open question.

In the appendices we tie up some loose ends, including Appendix D, where we relate our results to some similar results given in [2, 14] with respect to more restricted sentences systems, which we call  $\mathcal{F}$ -systems.

## 1 A Functionally Complete Language of Paradox

For each set of sentence names  $\mathbb{S}$ , we will introduce an infinitary propositional language  $\mathcal{L}_{\mathbb{S}}$  which is functionally complete (i.e. expressively adequate in the sense that for every function  $g : \{0, 1\}^{\mathbb{S}} \rightarrow \{0, 1\}$  there is a sentence of  $\mathcal{L}_{\mathbb{S}}$  that expresses  $g$ ). We will then endow  $\mathcal{L}_{\mathbb{S}}$  with a reference structure by adding a layer of arbitrary denotation relations between the sentence names (i.e. the proposition letters) and the formulae of  $\mathcal{L}_{\mathbb{S}}$ . These propositional languages endowed with denotation relations will provide all the complexity needed for

---

<sup>4</sup>See [6, p. 150] and [7, p. 693].

our general investigation into the nature of paradox (hypodox) supporting reference structures.

### 1.1 Syntax for $\mathcal{L}_S$

For a set of *sentence names*  $S$  (of arbitrary cardinality) we define a language  $\mathcal{L}_S$  as follows.  $\mathcal{L}_S$  contains the sentence names  $S$ , the nullary operators  $\top$  and  $\perp$ , the unary operator  $\neg$ , a binary operator  $\wedge$  and the operator  $\bigwedge$ . The collection  $S^+$  of well-formed *sentences* of  $\mathcal{L}_S$  is given by the following definition.

- Both  $\top$  and  $\perp$  are sentences.
- For each  $\alpha \in S$ ,  $\alpha$  is a sentence.
- If  $\phi$  is a sentence, then  $\neg\phi$  is a sentence.
- If  $\phi$  and  $\psi$  are sentences, then  $\phi \wedge \psi$  is a sentence.
- If  $I$  is an infinite set and  $\{\phi_i\}_{i \in I}$  is a sequence of sentences, then  $\bigwedge_{i \in I} \phi_i$  is a sentence.<sup>5</sup>
- Nothing else is a sentence.

Notice that we have defined a language  $\mathcal{L}_S$  for any given set of sentence names  $S$ . We will often speak as if there is one language  $\mathcal{L}_S$  but bear in mind that we are really talking about every language  $\mathcal{L}_S$ , unless otherwise stated.

We also define some shorthand for the language to ease our exposition. For sentences  $\phi$  and  $\psi$  in  $S^+$ , let  $\phi \vee \psi$  be shorthand for the sentence  $\neg(\neg\phi \wedge \neg\psi)$ . Additionally, if  $I$  is an infinite set and  $\{\phi_i\}_{i \in I}$  is a sequence of sentences, we write  $\bigvee_{i \in I} \phi_i$  as shorthand for  $\neg \bigwedge_{i \in I} \neg\phi_i$ .

### 1.2 Semantics for $\mathcal{L}_S$

It will help our exposition here and throughout the remainder of the paper, if we setup a way of speaking in the metalanguage which mirrors our language  $\mathcal{L}_S$ .<sup>6</sup> So let's introduce operations on the model-theoretic domain for our language by bestowing  $\{0, 1\}$  with the usual boolean algebra structure. That is, for  $x, y \in \{0, 1\}$ ,

- $\neg x = \begin{cases} 0 & \text{if } x = 1 \\ 1 & \text{if } x = 0 \end{cases}$ ,
- $x \wedge y = \begin{cases} 1 & \text{if } x = 1 \text{ and } y = 1 \\ 0 & \text{otherwise} \end{cases}$ ,

<sup>5</sup>Our proof of functional completeness will show that we would still get a functionally complete language if we placed the restriction  $|I| \leq 2^{|S|}$  on  $I$ .

<sup>6</sup>In particular, by using the same symbols in the object and metalanguage, manipulations are easier to follow since the basic equivalences that are true *look like they should be true*. We do respect the distinction between the metalanguage and the object language throughout, but we've tried to not be overly pedantic. Our hope is that this strategy reduces the cognitive work for the reader, without causing undue confusion. Additionally, in our judgment this convention makes the presentation much more aesthetically pleasing.

$$- \quad x \vee y = \neg(\neg x \wedge \neg y).$$

Additionally, for any set  $I$  and sequence  $\{x_i\}_{i \in I}$  with  $x_i \in \{0, 1\}$  we let  $\bigwedge_{i \in I} x_i$  be 1 if each  $x_i$  is 1 and zero otherwise. Similarly to  $x \vee y$ , we put  $\bigvee_{i \in I} x_i = \neg \bigwedge_{i \in I} \neg x_i$ .

Now we define the compositionally determined truth-value of any sentence of  $\mathcal{L}_S$  relative to an interpretation of the sentence names. Let a *truth-value assignment* be a function  $v$  from the sentence names  $S$  to  $\{0, 1\}$ . Then for all  $\chi \in S^+$  we define  $\llbracket \chi \rrbracket(v)$  for a truth-value assignment  $v$  as follows:<sup>7</sup>

- $\llbracket \top \rrbracket(v) = 1,$
- $\llbracket \perp \rrbracket(v) = 0,$
- For all  $\alpha \in S, \llbracket \alpha \rrbracket(v) = v(\alpha),$
- For all  $\phi \in S^+, \llbracket \neg \phi \rrbracket(v) = \neg \llbracket \phi \rrbracket(v),$
- For all  $\phi, \psi \in S^+, \llbracket \phi \wedge \psi \rrbracket(v) = \llbracket \phi \rrbracket(v) \wedge \llbracket \psi \rrbracket(v).$
- If  $I$  is an infinite set and  $\{\phi_i\}_{i \in I}$  is a sequence of sentences  $\llbracket \bigwedge_{i \in I} \phi_i \rrbracket(v) = \bigwedge_{i \in I} \llbracket \phi_i \rrbracket(v).$

Let  $V_S$  be the set of all truth-value assignments on  $S$ . For  $\chi \in S^+,$  we write  $\llbracket \chi \rrbracket$  for  $\chi$ 's associated function from  $V_S$  to  $\{0, 1\}$  (we also call this the function induced by  $\chi$ ).

### 1.3 Functional Completeness of $\mathcal{L}_S$

$\mathcal{L}_S$  lacks both truth and falsity predicates and first-order quantification but, in the relevant sense, there is nothing that cannot be expressed in the language. To see that adding the truth and falsity predicates (say  $\top$  and  $\perp$ ) would not increase expressive power, note that in the pairs  $(\top(\alpha), \alpha)$  and  $(\perp(\alpha), \neg \alpha)$  the sentences induce the same function from truth-value assignments to  $\{0, 1\}$  as their respective pair-mate. Given that the size of the set of sentence names  $S$  can have arbitrary cardinality and that there is no restriction on the length of sentences in  $S^+,$  the addition of first-order quantifiers would also not add expressive power. Those are intuitive reasons why our expressive power would not be increased by the addition of these familiar devices. Now we prove that  $\mathcal{L}_S$  is in fact expressively adequate.

**Lemma 1** *For any function  $g$  from  $V_S$  to  $\{0, 1\}$  we have a sentence  $\zeta_g \in S^+$  such that  $\llbracket \zeta_g \rrbracket = g.$*

*Proof* Let  $g$  be a function from  $V_S$  to  $\{0, 1\}$ . First, if  $g$  is a constant function, put  $\zeta_g = \top$  if  $g$  maps everything to 1 and  $\zeta_g = \perp$  if  $g$  maps everything to 0. Otherwise, the strategy is to first decompose  $g$  using Kronecker's  $\delta$  function and then exhibit subsentences which induce the simple parts of  $g$ . From these subsentences we then construct the desired complex sentence. Recall

---

<sup>7</sup>Read " $\llbracket \chi \rrbracket(v)$ " as "the truth-value of  $\chi$  relative to assignment  $v$ ".

that Kronecker's  $\delta$  is a function of two arguments (in this case two truth-assignments), which outputs 1 if the arguments are identical and 0 otherwise. So for  $r, v \in \mathcal{V}_S$ ,

$$\delta_{rv} = \begin{cases} 1 & \text{if } r = v \\ 0 & \text{if } r \neq v \end{cases}.$$

Notice that for every  $r \in \mathcal{V}_S$

$$g(r) = \bigvee_{v \in \mathcal{V}_S} \delta_{rv} g(v).$$

since for every  $v$  distinct from  $r$ ,  $\delta_{rv} g(v) = 0$  and when  $v$  just is  $r$ ,  $\delta_{rr} g(r) = 1g(r)$  and  $g(r)$  plus a bunch of 0's still equals  $g(r)$ . In the equation above  $g(v)$  only makes a difference when it equals 1, so we can restrict our focus to the  $v$ 's where  $g(v) = 1$ . Let  $C = \{v \in \mathcal{V}_S \mid g(v) = 1\}$ . Now for each  $r \in \mathcal{V}_S$ , we see that

$$g(r) = \bigvee_{v \in C} \delta_{rv}.$$

Thus it will be sufficient to construct, for each  $v \in C$ , a sentence  $\chi_v$  such that  $\llbracket \chi_v \rrbracket(r) = \delta_{rv}$  for each  $r \in \mathcal{V}_S$ .

For  $v \in \mathcal{V}_S$  and  $\alpha \in \mathbf{S}$ , let

$$h(v, \alpha) = \begin{cases} \alpha & \text{if } v(\alpha) = 1 \\ \neg\alpha & \text{if } v(\alpha) = 0 \end{cases}.$$

Then, for  $v \in C$ , define

$$\chi_v = \bigwedge_{\alpha \in \mathbf{S}} h(v, \alpha).$$

Note that for every  $v \in C$  and for every  $r \in \mathcal{V}_S$ ,  $\llbracket \chi_v \rrbracket(r) = \delta_{rv}$ , since by design  $\llbracket \chi_v \rrbracket(r) = 1$  iff  $r = v$ . We have already established that for all  $r \in \mathcal{V}_S$

$$g(r) = \bigvee_{v \in C} \delta_{rv},$$

so it follows that for all  $r \in \mathcal{V}_S$

$$g(r) = \bigvee_{v \in C} \llbracket \chi_v \rrbracket(r).$$

Thus we may let  $\zeta_g$  be the sentence

$$\bigvee_{v \in C} \chi_v.$$

□

#### 1.4 Denotation Assignments and Paradoxicality

Thus far we have defined an expressively adequate infinitary propositional language  $\mathcal{L}_S$ . But clearly something is lacking, since nothing yet models

the notions of “reference”, in the sense of “self-reference”.<sup>8</sup> So we cannot construct paradoxes in  $\mathcal{L}_S$ . Consider the normal statement of the *liar sentence*:

(1) This sentence is not true.

Sentence (1) “references” itself since the complex demonstrative ‘this sentence’ contained therein refers to sentence (1).<sup>9</sup> We have no such resources in  $\mathcal{L}_S$ . We have formulae like  $\neg\alpha$  but there is no way to *link*  $\alpha$  with  $\neg\alpha$ . Another common way to state the liar paradox is as follows.

$L$ :  $L$  is not true.

The colon is to be read as “refers to” or “denotes”, so that ‘ $L$ ’ denotes the sentence ‘ $L$  is not true’—just as ‘this sentence’ denotes sentence (1) above. If ‘ $L$ ’ denotes the sentence ‘ $L$  is not true’, then  $L = \text{‘}L \text{ is not true’}$ —compare: if ‘Cicero’ denotes Tully, then Cicero = Tully. This, then seems to be the natural language phenomenon we need to model. We need to define a relation on  $\mathcal{L}_S$  between the sentence names  $S$  and the sentences  $S^+$ . And the relation should, in fact, be a function, since we do not want sentence names to denote multiple sentences. Let’s call this a “denotation” assignment, even though it needn’t be thought of in terms of the members of  $S$  *denoting* sentences—it is just an arbitrary mapping from  $S$  to  $S^+$ , which models the natural language phenomenon.<sup>10</sup>

<sup>8</sup> For a paradigmatic example of self-reference consult [9], footnote 8.

<sup>9</sup>We should flag that the phrase “what a sentence refers to” is often used in two distinct ways. The sense in which the Liar sentence “refers” to itself should not be confused with claims about what the *reference* of the Liar sentence is, in the sense of its “extension” in a truth-conditional semantics. The Liar sentence “refers” to itself due to it containing a referential device (name, demonstrative, etc.) which refers to (designates) the Liar sentence. Everything we say here is consistent with the Fregean doctrine that sentences refer to (designate) their truth-values—the sentence ‘ $A \wedge \neg A$ ’ *refers* to das Falsche but *references*  $A$ .

<sup>10</sup>Given the way we have set things up the elements of  $S$  play two important roles that may seem to be in conflict: (i) they are genuine sentences (and arguments to the truth-value assignment  $v$ ) and (ii) they are “names” of sentences, in the sense that the denotation function  $d$  maps them to sentences in  $S^+$ . If the atomic sentences (i.e. the elements of  $S$ ) “denote” sentences, then we might expect that they should “denote” themselves. But given our setup an atomic sentence not only can fail to “denote” itself it can “denote” any sentence whatsoever (including its own negation).

The framework in [2] has the potential conceptual advantage in that the expressions that do the “denoting” are not genuine sentences of the formal language—they are simply *names* of sentences that can occur as syntactic constituents of sentences. So if a sentence name  $\alpha$  is a name of a *sentence* but  $\alpha$  itself isn’t a sentence then we should not expect that it denote itself.

This, however, introduces a conceptual oddity of its own. On this setup sentence names will also play two important roles: (i) they are names of sentences and (ii) they are “truth-evaluable” expressions. One might think that if a sentence name is not a *genuine sentence* then it is not truth-evaluable—names are not true or false.

We could make our presentation more in line with the strategy of [2] without endangering any of the results. We would simply need to introduce a truth predicate  $T$  and have our atomic sentences be of the form  $\ulcorner T(\alpha) \urcorner$ , for all  $\alpha \in S$ . We would thereby downgrade the elements of  $S$  from their sentential status. And the necessary equivalence in truth-value between every  $\alpha \in S$  and its corresponding sentence  $T(\alpha)$  would ensure that nothing of mathematical import is impacted (although the proofs would become significantly more complicated and messy). But given the equivalence we find it desirable to use the simplified formulation.

**Definition 1** A denotation assignment is a function  $d$  from  $\mathbf{S}$  to  $\mathbf{S}^+$ .

There is another feature of “reference”, which we must also account for. If ‘Atticus’ refers to the sentence ‘Aardvarks swim’, then Atticus is a true sentence if and only if ‘Aardvarks swim’ is a true sentence. The liar paradox rests on this type of inference, e.g. when we assume that  $L$  is true and then infer that the sentence ‘ $L$  is not true’ is true, this is justified by the assumption that ‘ $L$ ’ denotes ‘ $L$  is not true’. In general, then, our denotation assignment constrains which truth-value assignments are acceptable—the only acceptable assignments are the ones that assign to a sentence name a value identical to the truth-value of the sentence it denotes.

**Definition 2** A truth-value assignment  $v$  is acceptable on  $\mathbf{S}$  relative to  $d$  if and only if for every  $\alpha \in \mathbf{S}$ ,  $v(\alpha) = \llbracket d(\alpha) \rrbracket(v)$ .

With a denotation assignment layered on top of our language  $\mathcal{L}_{\mathbf{S}}$ , we can construct paradoxes. For example, if we let  $\mathbf{S} = \{L\}$  and  $d(L) = \neg L$ , then for any acceptable truth assignment  $v$  we have

$$v(L) = \llbracket \neg L \rrbracket(v) = \neg \llbracket L \rrbracket(v) = \neg v(L).$$

This is a contradiction.  $L$  relative to  $d$  is our formal representation of the *liar paradox*.<sup>11</sup> Notice that here the things that are paradoxical are not just sentences of  $\mathcal{L}_{\mathbf{S}}$  but sets of sentences of  $\mathcal{L}_{\mathbf{S}}$  relative to a denotation assignment—we call these pairs  $(\mathbf{S}, d)$  *sentence systems*.

**Definition 3** The sentence system  $(\mathbf{S}, d)$  is *paradoxical* if there is no acceptable truth-value assignment on  $\mathbf{S}$  relative to  $d$ .

For another example, consider *Yablo’s paradox*.<sup>12</sup> Let  $\mathbf{S} = \{Y_1, Y_2, Y_3, \dots\}$  and for each  $Y_k \in \mathbf{S}$ , let  $d(Y_k) = \bigwedge_{j>k} \neg Y_j$ . Then every sentence  $Y_k$  says that every subsequent sentence is false:

$$\begin{aligned} d(Y_1) &= \neg Y_2 \wedge \neg Y_3 \wedge \neg Y_4 \wedge \dots \\ d(Y_2) &= \neg Y_3 \wedge \neg Y_4 \wedge \neg Y_5 \wedge \dots \\ d(Y_3) &= \neg Y_4 \wedge \neg Y_5 \wedge \neg Y_6 \wedge \dots \\ &\vdots \end{aligned}$$

<sup>11</sup>Whereas the common representation of the liar is “ $L$ :  $L$  is not true”, we have “ $d(L) = \neg L$ ”. It should be noted that our denotation function  $d$  plays an analogous role to the relation given by “:” in the more common formulations.

<sup>12</sup>See [13].



The set of sentences  $\{Y_1, Y_2, Y_2, \dots\}$  are paradoxical relative to  $d$ . If  $v$  is an acceptable truth assignment, then

$$v(Y_k) = \llbracket \bigwedge_{j>k} \neg Y_j \rrbracket(v) = \bigwedge_{j>k} \llbracket \neg Y_j \rrbracket(v) = \bigwedge_{j>k} \neg \llbracket Y_j \rrbracket(v) = \bigwedge_{j>k} \neg v(Y_j).$$

In particular, for each  $k$ ,

$$v(Y_k) = \neg v(Y_{k+1}) \wedge \bigwedge_{j>k+1} \neg v(Y_j) = \neg v(Y_{k+1}) \wedge v(Y_{k+1}) = 0.$$

Thus,  $0 = v(Y_0) = \bigwedge_{j>0} \neg v(Y_j) = \bigwedge_{j>0} \neg 0 = 1$ . A contradiction.

There is another class of self-referential puzzles, which don't come out as paradoxical on these definitions. Consider the sentence that says of itself that it is true (i.e. the *truth-teller*). We represent this as the sentence system  $S = \{T\}$  and  $d(T) = T$ . For any truth assignment  $v$  we have

$$\llbracket T \rrbracket(v) = v(T).$$

So any truth assignment is acceptable. But is the correct assignment  $v(T) = 1$  or  $v(T) = 0$ ? The problem is that nothing decides one truth assignment over the other. Unlike a paradox where we are pulled in both directions, here we are pulled in *neither* direction. We will call such situations *hypodoxical*, since here the truth-value is underdetermined.<sup>13</sup>

In the case of the truth-teller, the reference structure employed (i.e. self-reference), was the same as the one employed in the liar paradox. But it could be that the reference structures that support paradoxes are fundamentally different from those that support hypodoxes, so it is desirable to pull these notions apart.

**Definition 4** The sentence system  $(S, d)$  is *hypodoxical* if there is more than one acceptable truth-value assignment on  $S$  relative to  $d$ .

### 1.5 Duals of Paradox and Hypodox

Cook [2] demonstrates how to turn a paradox given in terms of conjunction and a falsity predicate into a paradox given in terms of disjunction and a falsity predicate.<sup>14</sup> We provide a direct generalization of this for our more expressive languages. Anytime we have a paradox (hypodox) we can construct a *dual* of the paradox (hypodox) with isomorphic reference relations. To define this notion of dual precisely we first need to introduce notation for substitution.

<sup>13</sup>We adopt the term “hypodox” from [5]. Cook [2] uses the term “indeterminate” for the analogous notion with respect to  $\mathcal{F}$ -systems.

<sup>14</sup>See [2, pp. 771–772]. Cook limits his focus to languages with conjunction and a falsity predicate (i.e. what we call  $\mathcal{F}$ -systems; see Appendix D).

**Definition 5** Let  $\mathbf{S}$  be a set of names and let  $\psi \in \mathbf{S}^+$ . Let  $\{(\alpha_i, \gamma_i)\}_{i \in I} \subseteq \mathbf{S} \times \mathbf{S}^+$  such that the  $\alpha_i$  are pairwise distinct. We write  $\psi [\alpha_i \Rightarrow \gamma_i \mid i \in I]$  for the sentence obtained from  $\psi$  by replacing each  $\alpha_i$  with  $\gamma_i$ .<sup>15</sup>

For example, for  $\{(A_1, B_1 \wedge B_2), (A_2, \neg A_2)\} \subseteq \mathbf{S} \times \mathbf{S}^+$ ,  $(A_1 \wedge (\neg A_1 \vee A_2)) [A_i \Rightarrow B_i \mid i \in I] = ((B_1 \wedge B_2) \wedge (\neg(B_1 \wedge B_2) \vee \neg A_2))$ .

**Definition 6** For a sentence system  $(\mathbf{S}, d)$  the *dual* denotation assignment  $d^*$  on  $\mathbf{S}$  is given by  $d^*(\alpha) = \neg d(\alpha)[\beta \Rightarrow \neg \beta \mid \beta \in \mathbf{S}]$ .

It is not difficult to check that  $v$  is an acceptable truth assignment on  $\mathbf{S}$  with respect to  $d$  if and only if  $v^*$  defined by  $v^*(\alpha) = \neg v(\alpha)$  is an acceptable truth assignment on  $\mathbf{S}$  with respect to  $d^*$ .

For an easy example consider *Jourdain's paradox*. Let  $\mathbf{S} = \{J_1, J_2\}$  and let  $d$  be such that  $d(J_1) = \neg J_2$  and  $d(J_2) = J_1$ . So  $J_1$  says that  $J_2$  is false and  $J_2$  says that  $J_1$  is true. There is no acceptable truth assignment for the system  $(\mathbf{S}, d)$ , since for any acceptable truth assignment  $v$  we have

$$v(J_2) = \llbracket J_2 \rrbracket(v) = v(J_1) = \llbracket \neg J_2 \rrbracket(v) = \neg \llbracket J_2 \rrbracket(v) = \neg v(J_2).$$

Thus,

$$v(J_2) = \neg v(J_2).$$

Then by Definition 6 the dual of Jourdain's paradox is given by the following:

$$d^*(J_1) = \neg d(J_1)[\beta \Rightarrow \neg \beta \mid \beta \in \mathbf{S}] = \neg(\neg \neg J_2),$$

$$d^*(J_2) = \neg d(J_2)[\beta \Rightarrow \neg \beta \mid \beta \in \mathbf{S}] = \neg(\neg J_1).$$

We see that there is also no acceptable truth assignment for the system  $(\mathbf{S}, d^*)$ , since for any acceptable truth assignment  $v$  we have

$$v(J_2) = \llbracket \neg(\neg J_1) \rrbracket(v) = v(J_1) = \llbracket \neg(\neg \neg J_2) \rrbracket(v) = \neg \llbracket J_2 \rrbracket(v) = \neg v(J_2).$$

And again,

$$v(J_2) = \neg v(J_2).$$

## 2 Reference Graphs and Danger

With the addition of denotation assignments we now have the ability to represent paradoxes in  $\mathcal{L}_{\mathcal{S}}$  by sentence systems  $(\mathbf{S}, d)$ . Since our overall aim is to study the “reference structures” that support paradoxes, we need a way of extracting reference structural information from sentence systems. Looking at an example provides guidance on how to go about extracting this information.

<sup>15</sup>We will abuse this notation slightly in places by writing  $d(\alpha)[\beta \Rightarrow w]$  to represent replacing all occurrence of  $\beta$  with  $w$  even when  $w$  is not in  $\mathbf{S}^+$ .



**Fig. 1** The reference structure of Jourdain’s paradox

Consider again Jourdain’s paradox, where  $S = \{J_1, J_2\}$  and  $d$  is such that  $d(J_1) = \neg J_2$  and  $d(J_2) = J_1$ . What is the reference structure supporting this paradox? Intuitively,  $J_1$  references  $J_2$  and  $J_2$  references  $J_1$ , so that there is a pattern of cyclic reference (Fig. 1).

As a first pass at defining the reference structure we might simply use the inputs and outputs of the denotation assignment. But we immediately see that this is incorrect. On this approach there would be no sense in which  $J_1$  references  $J_2$ —it would only reference  $\neg J_2$ . And if  $J_1$  referenced  $\neg J_2$ , then there would be no cyclic reference, since the denotation assignment doesn’t map  $\neg J_2$  anywhere. Instead we should say that it is in virtue of the denotation assignment mapping  $J_1$  to the sentence  $\neg J_2$ , that  $J_1$  references  $J_2$ . This motivates the following definition.

**Definition 7** We say that a sentence name  $\alpha \in S$  *references* a sentence name  $\beta \in S$  with respect to a denotation assignment  $d$  if the name  $\beta$  occurs as a syntactic constituent of  $d(\alpha)$ .

With this definition of *referencing* in play the “reference structure” of any given sentence system  $(S, d)$  is best encoded as a directed graph in the following manner.<sup>16</sup>

**Definition 8** Let  $S$  be a set of sentence names and  $d$  a denotation assignment. The *reference graph*  $G_{S,d}$  is the directed graph with vertex set  $S$  and an edge from  $\alpha \in S$  to  $\beta \in S$  if and only if  $\alpha$  references  $\beta$ .

For example, we can now represent the reference graph of the Liar paradox as the self-loop, since  $d(L) = \neg L$  and  $L$  occurs as a syntactic constituent of  $\neg L$  (Fig. 2).

In Yablo’s paradox every sentence  $Y_i$  denotes a sentence which has all the sentences names  $Y_j$ , for  $j > i$ , as syntactic constituents. So we get the result that in the reference graph every sentence name references every “later” sentence name (Fig. 3).

<sup>16</sup>A directed graph is just a binary *relation* on a certain domain, i.e. a directed graph  $G$  is a pair  $(V, E)$  where  $E$  is a set of ordered pairs of elements from  $V$ . For the basics of graph theory see [4]. We will do our best to not assume that the reader is familiar with graph-theoretic terminology but a quick review of the basics may prove helpful.

**Fig. 2** The Liar graph

We'd like to provide necessary and sufficient conditions for paradox (hypodox) supporting reference graphs. For a reference graph to “support” a paradox (hypodox) we mean that there is at least one paradoxical (hypodoxical) sentence system  $(S, d)$  with that reference structure. So the self-loop is paradox and hypodox supporting since the reference graph of both the Liar paradox and the Truth-teller hypodox are isomorphic to the self-loop. This notion of paradox and hypodox supporting graphs is the central concern of this essay. We call paradox supporting graphs *dangerous* and hypodox supporting graphs *precarious*.

**Definition 9** We call a directed graph  $G$  *dangerous* if there exists a paradoxical sentence system  $(S, d)$  such that  $G$  is isomorphic to  $\mathcal{G}_{S,d}$ .

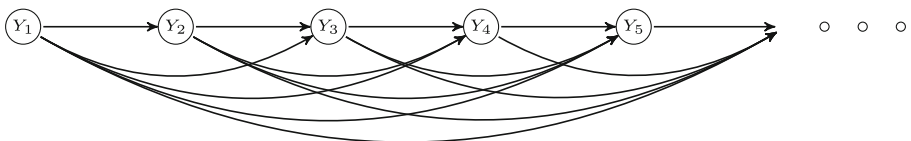
**Definition 10** We call a directed graph  $G$  *precarious* if there exists a hypodoxical sentence system  $(S, d)$  such that  $G$  is isomorphic to  $\mathcal{G}_{S,d}$ .

The problem that guides our investigation, then, is this.

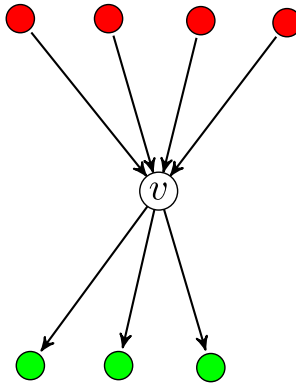
**Problem 1** Classify the dangerous (precarious) directed graphs.

### 3 Danger Preserving Operations

As a first theoretical step toward solving this problem it is useful to find some basic graph operations that preserve danger (precarity). Having these operations at our disposal allows us to separate the wheat from the chaff by reducing complex graphs to simpler—yet still dangerous (precarious)—graphs. In this way we can identify the salient properties and gain a better understanding of the space of dangerous (precarious) graphs. We will make use of these operations throughout the essay to help generate new paradoxes from old ones, find counterexamples to conjectures, and for proving the later theorems.

**Fig. 3** The Yablo graph

**Fig. 4**  $N_G^-(v)$  in red and  $N_G^+(v)$  in green



### 3.1 Important Graph-Theoretic Notions

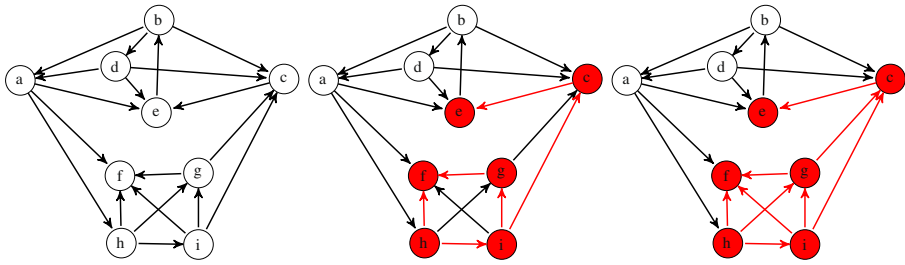
Here we must introduce some essential graph-theoretic notions. One important concept is that of a vertex’s *neighbors*. In a directed graph each vertex has both (i) an associated set of vertices, which are the vertices it points to, (ii) and an associated set of vertices, which are the vertices that point to it. We call these a vertex’s *out-neighbors* and *in-neighbors*, respectively.

**Definition 11** Let  $G$  be a directed graph. We write  $N_G^+(v)$  for or the set containing exactly the vertices  $x$  in  $G$  such that there is an edge from  $v$  to  $x$  ( $v$ ’s out-neighbors) and  $N_G^-(v)$  for the set containing exactly the vertices  $x$  in  $G$  such that there is an edge from  $x$  to  $v$  ( $v$ ’s in-neighbors) (Fig. 4).

Another important notion is that of a *subgraph*. Intuitively, a subgraph of a graph is some “part” of the graph. But there are different ways to take a part of a graph. You can either take a subset of the vertex set and thrown out some of the edges therein or take a subset of the vertex set and retain all the edges therein. The first is the general notion of a subgraph (i.e. part) and the latter is the more specific concept of an *induced* subgraph (i.e. whole part). More precisely these are defined as follows.

**Definition 12** Let  $G$  be a directed graph. A *subgraph*  $H$  of  $G$  is a directed graph with vertex set  $V(H)$  a subset of  $V(G)$  and edge set  $E(H)$  a subset of  $E(G)$  such that if  $xy \in E(H)$ , then  $x, y \in H$ . An *induced subgraph*  $H$  of  $G$  is a subgraph of  $G$  such that if  $x, y \in H$  and  $xy \in E(G)$ , then  $xy \in E(H)$ . For  $A \subseteq V(G)$  we write  $G[A]$  for the induced subgraph of  $G$  with vertex set  $A$  (Fig. 5).

Now we turn to the danger preserving operations.



**Fig. 5** A directed graph  $G$ , a subgraph of  $G$  with vertex set  $\{c, e, f, g, h, i\}$  and the graph  $G[c, e, f, g, h, i]$

### 3.2 Subgraphs

It seems that a graph should be dangerous (precarious) if and only if some part of it is. We can prove that this is indeed that case. The forward implication is obvious and the reverse implication is proved by assuming that a subgraph of a graph is dangerous (precarious) and then demonstrating a procedure to construct an extended denotation assignment to the whole graph while preserving danger (precarity).

**Lemma 2** *Let  $G$  be a directed graph. Then  $G$  is dangerous (precarious) if and only if some subgraph of  $G$  is dangerous (precarious).*

*Proof* Since  $G$  is a subgraph of itself, the forward implication is immediate. For the reverse implication, let  $H$  be a subgraph of  $G$  that is dangerous (precarious). View  $V(H)$  as a set of sentence names and let  $d$  be a denotation assignment on  $V(H)$  such that  $(V(H), d)$  is paradoxical (hypodoxical) and  $H = \mathcal{G}_{V(H), d}$ .

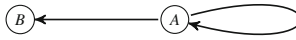
We construct a denotation assignment  $d'$  on  $V(G)$  by employing “junk conjunctions”  $J_x = \perp \wedge \bigwedge_{y \in N_G^+(x)} y$ , for each  $x \in V(G)$ . The denotation assignment  $d'$  assigns to  $x$  a sentence containing its junk conjunction—it is clear that this gets the referencing right, but the junk sentences must be added in such a way that they have no impact on the acceptability of truth-value assignments. For  $x \in V(G)$ , let

$$d'(x) = \begin{cases} d(x) \vee J_x & \text{if } x \in V(H) \\ J_x & \text{if } x \notin V(H) \end{cases}$$

Then, by construction,  $G = \mathcal{G}_{V(G), d'}$  and  $(V(G), d')$  is paradoxical (hypodoxical). Hence  $G$  is dangerous (precarious).  $\square$

For an example, consider the following version of Curry’s paradox. Put  $S = \{A, B\}$ ,  $d(A) = \neg A \vee B$  and  $d(B) = \perp$ . The reference graph of  $(S, d)$  is given in Fig. 6. By Lemma 2, we can determine that this graph is dangerous prior to considering Curry’s paradox, since it has the self-loop as a subgraph.

**Fig. 6** The Curry graph



Above we added junk conjunctions, which did the trick because they established the right reference relation while not affecting the acceptability of truth-assignments on the sentence system. They were able to do this because on every truth-assignment they got the same value, i.e. they were *constant*. We would also like to be able to remove constant junk from a sentence system, thereby eliminating certain reference relations without affecting the acceptability of truth-assignments.

For example, for the system  $S = A$  and  $d(A) = A \wedge \neg A$  we can modify  $d$  to  $d'(A) = \perp$  thereby removing the reference relations without affecting the acceptability of truth-value assignments. If this procedure is carried out for all the constant junk in a sentence system, we will say that the system is *junk-free*.<sup>17</sup> Let's define this notion and prove a lemma, while we are on the topic of "junk".

**Definition 13** A sentence system  $(S, d)$  is *junk-free* if and only if for every  $\alpha \in S$ , if  $d(\alpha) \notin \{\perp, \top\}$ , then there exist truth assignments  $v_0, v_1$  on  $S$  such that  $\llbracket d(\alpha) \rrbracket(v_0) = 0$  and  $\llbracket d(\alpha) \rrbracket(v_1) = 1$ .

Since removing junk does not affect the acceptability of truth-assignments on a system, the removal of junk preserves the paradoxicality (hypodoxicality) of a sentence system.

**Lemma 3** Let the system  $(S, d)$  be paradoxical (hypodoxical). Then there is a denotation assignment  $d'$  on  $S$  such that  $(S, d')$  is junk-free, paradoxical (hypodoxical) and  $\mathcal{G}_{S,d'}$  is a subgraph of  $\mathcal{G}_{S,d}$ .

*Proof* Assume  $(S, d)$  is paradoxical (hypodoxical). Let  $J$  be the set of  $\alpha \in S$  such that there is  $t_\alpha \in \{0, 1\}$  so that  $\llbracket d(\alpha) \rrbracket(v) = t_\alpha$  for each  $v \in V_S$ . For each  $\alpha \notin J$ , let  $d'(\alpha) = d(\alpha)$  and for  $\alpha \in J$ , let

$$d'(\alpha) = \begin{cases} \perp & \text{if } t_\alpha = 0 \\ \top & \text{if } t_\alpha = 1 \end{cases}$$

Then, by construction,  $(S, d')$  is junk-free, paradoxical (hypodoxical) and  $\mathcal{G}_{S,d'}$  is a subgraph of  $\mathcal{G}_{S,d}$ . □

### 3.3 Smoothing and Subdividing

For a sentence system  $(S, d)$ , intuitively, we can "simplify" it to the system  $(S', d')$  by picking some  $\alpha \in S$ , replacing every occurrence of  $\alpha$  with  $d(\alpha)$  and

<sup>17</sup>Note that a system that is junk-free may still include "junky disjuncts" such as in the right disjunct of  $B \vee (A \wedge \neg A)$ . Junk-free sentence systems are merely free of *constant* junk.

then removing  $\alpha$  from  $\mathbf{S}$ . Doing so gives rise to an additional danger preserving operation on graphs. This is supported by the following lemma.

**Lemma 4** *Let  $\mathbf{S}$  be a set and let  $\psi \in \mathbf{S}^+$ . Let  $v$  be a truth-value assignment on  $\mathbf{S}$ . Let  $\{(\alpha_i, \gamma_i)\}_{i \in I} \subseteq \mathbf{S} \times \mathbf{S}^+$  such that the  $\alpha_i$  are pairwise distinct. If  $v(\alpha_i) = \llbracket \gamma_i \rrbracket(v)$  for each  $i \in I$ , then*

$$\llbracket \psi [\alpha_i \Rightarrow \gamma_i \mid i \in I] \rrbracket (v) = \llbracket \psi \rrbracket (v).$$

*Proof* This is immediate from the definition of the semantics. □

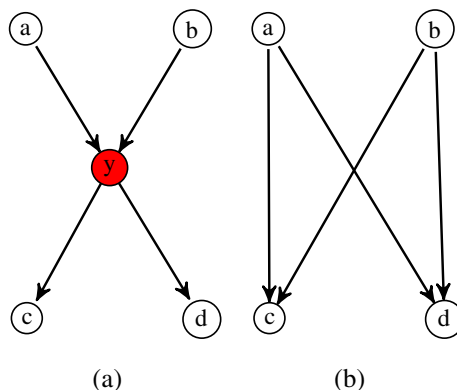
Now consider a pair  $(\mathbf{S}, d)$  and  $\alpha \in \mathbf{S}$  such that  $\alpha$  does not occur as a constituent of  $d(\alpha)$ . Let  $\mathbf{S}' = \mathbf{S} - \{\alpha\}$  and for  $\beta \in \mathbf{S}'$  let  $d'(\beta) = d(\beta) [\alpha \Rightarrow d(\alpha)]$ . Then  $d'$  is a denotation assignment on  $\mathbf{S}'$ . Given a truth assignment  $v$  on  $\mathbf{S}$ , let  $v'$  be  $v$  restricted to  $\mathbf{S}'$ . Then, by Lemma 4,  $v$  is acceptable on  $\mathbf{S}$  relative to  $d$  if and only if  $v'$  is acceptable on  $\mathbf{S}'$  relative to  $d'$ . Thus the following “smoothing” operation leaves dangerous graphs dangerous and precarious graphs precarious.

**Definition 14** Let  $G$  be a directed graph. Let  $y \in V(G)$  such that  $yy \notin E(G)$ . The *smoothing* of  $G$  at  $y$  is the graph  $H$  with  $V(H) = V(G) - y$  and  $E(H) = E(G - y) \cup \{ab \mid a \in N^-(y), b \in N^+(y)\}$  (Fig. 7).

The reverse operation of introducing a new name doesn’t behave nicely in general, but it does in a few interesting special cases. One is that of a subdivision.

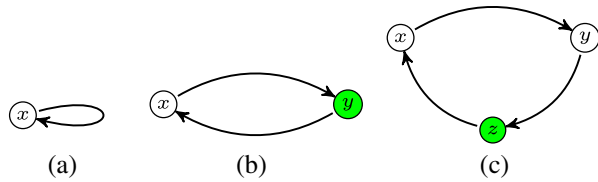
**Definition 15** Let  $G$  be a directed graph. A *subdivision* of  $G$  is a graph formed by replacing each edge  $xy$  of  $G$  with a path  $p_{xy}$  from  $x$  to  $y$ . Note that we allow  $p_{xy}$  to be length 1; that is,  $p_{xy} = xy$  (Fig. 8).

**Fig. 7** A graph and its smoothing at  $y$





**Fig. 8** The Liar graph and two subdivisions



**Lemma 5** *A subdivision of a directed graph  $G$  is dangerous if and only if  $G$  is.*<sup>18</sup>

*Proof* First, assume  $G$  is dangerous and let  $d$  be a denotation assignment on  $V(G)$  such that  $G = \mathcal{G}_{V(G),d}$  and  $(V(G), d)$  is paradoxical. Let  $H$  be the subdivision of  $G$  where each edge  $xy \in E(G)$  is replaced with the path  $p_{xy}$  from  $x$  to  $y$ . Say the vertices of  $p_{xy}$  in order are  $x = z_{xy}^1, z_{xy}^2, \dots, z_{xy}^{k_{xy}} = y$ . Define a denotation assignment  $d'$  on  $V(H)$  by

$$d'(w) = \begin{cases} d(w)[y \Rightarrow z_{wy}^2 \mid wy \in E(G)] & \text{if } w \in V(G) \\ z_{xy}^{j+1} & \text{if } w = z_{xy}^j \text{ for } xy \in E(G), j \geq 2. \end{cases}$$

By construction we have  $H = \mathcal{G}_{V(H),d'}$ . Now we will show that  $H$  is paradoxical with this denotation assignment. Assume not and let  $v'$  be an acceptable truth assignment on  $V(H)$  with respect to  $d'$  and let  $v$  be  $v'$  restricted to  $V(G)$ . Then for each  $xy \in E(G)$  and  $2 \leq j < k_{xy}$  we have  $\llbracket d'(z_{xy}^j) \rrbracket(v') = \llbracket z_{xy}^{j+1} \rrbracket(v')$ . Hence  $v'(z_{xy}^2) = \llbracket d'(z_{xy}^2) \rrbracket(v') = \llbracket d'(y) \rrbracket(v') = v'(y)$  since  $v'$  is acceptable. Thus for  $w \in V(G)$ , by Lemma 4, we have  $v(w) = v'(w) = \llbracket d'(w) \rrbracket(v') = \llbracket d(w)[y \Rightarrow z_{wy}^2 \mid wy \in E(G)] \rrbracket(v') = \llbracket d(w) \rrbracket(v') = \llbracket d(w) \rrbracket(v)$ . Whence  $v$  is acceptable on  $V(G)$  with respect to  $d$ . This contradicts the fact that  $G$  is dangerous.

The other direction is very similar, basically we smooth at all of the subdivision vertices. We omit the proof to avoid unnecessary tedium.  $\square$

We should mention here that the fact that subdivision preserves danger entails that danger is preserved under homeomorphisms and thus that being dangerous is a *topological property* of directed graphs. The terms “homeomorphism” and “topological property” are usually applied to topological spaces and in particular undirected graphs. Even though a directed graph is not a topological space, the straightforward generalization of the definitions gives a useful concept.

**Definition 16** Two directed graphs  $G$  and  $H$  are *homeomorphic* if some subdivision of  $G$  is isomorphic to some subdivision of  $H$ .

<sup>18</sup>This also holds for precarious graphs but we omit the proof here as it turns out to be quite nasty to write down—and in any case follows trivially from the later results about precariousity and groundedness (see Theorem 11). Similar remarks apply to precariousity’s status as a topological property.

**Lemma 6** *If  $G$  and  $H$  are homeomorphic directed graphs, then  $G$  is dangerous if and only if  $H$  is dangerous.*

*Proof* Assume  $G$  and  $H$  are homeomorphic directed graphs and let  $G'$  and  $H'$  be subdivisions of  $G$  and  $H$  respectively such that  $G'$  is isomorphic to  $H'$ . The lemma follows by applying Lemma 5 to  $G$  and  $G'$  and to  $H$  and  $H'$ .  $\square$

### 3.4 Unwinding

We give a construction that turns any finite paradoxical sentence system into an infinite acyclic paradoxical sentence system.<sup>19</sup> Following [2], we call this construction *unwinding the paradox* (see Fig. 9).<sup>20</sup> Cook gives a nice description of the philosophical import of unwinding.

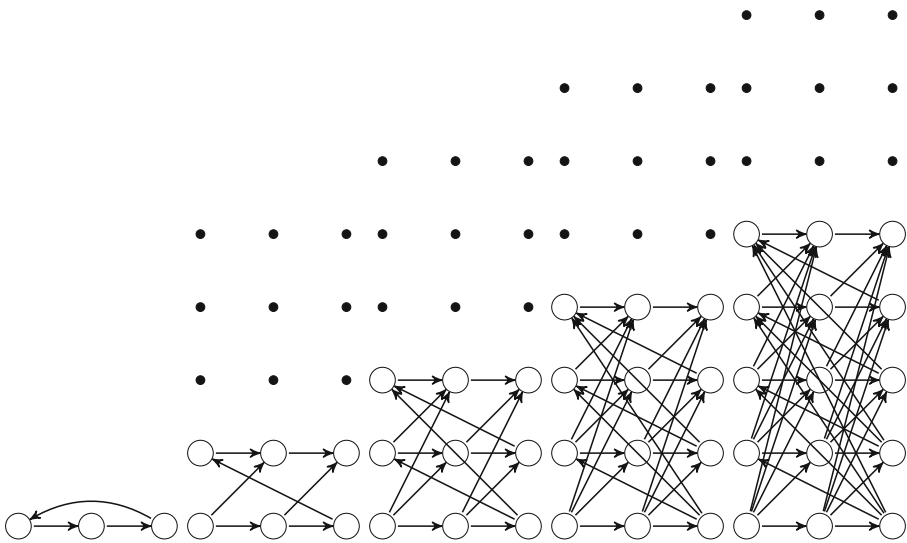
The notion of unwinding sheds considerable light onto the relevance of Yablo's paradox to debates regarding the connections between paradox and circularity. Prior to Yablo's discovery, semantic paradox was thought by most to be inextricably linked to circularity. Post Yablo, however, it is fair to say that most philosophers of language and logic think of the infinite non-circular construction as an isolated curiosity. Unwinding finite paradoxes, however, demonstrates that infinite non-circular constructions can be associated with each instance of a wide class of circular finite paradoxes. As a result, we might be forced to reconsider the assumption that circularity has any fundamental connection to the logical and philosophical problems associated with truth. [2, p. 772]

In what follows we show that an infinite non-circular construction can be associated with each finite paradox. More precisely, we show that from any finite paradoxical (hypodoxical) sentence system we can get an infinite *acyclic* paradoxical (hypodoxical) sentence system that is paradoxical (hypodoxical) for essentially the same reason.

**Definition 17** Let  $G$  be a finite directed graph and let  $<$  be a total order on  $V(G)$ . Define an ordering on  $V(G) \times \mathbb{N}$  by  $(x_1, i) < (x_2, j)$  if and only if  $i < j$  or  $i = j$  and  $x_1 < x_2$ . The *unwinding* of  $G$  is the graph with vertex set  $V(G) \times \mathbb{N}$  with an edge from  $(x_1, i)$  to  $(x_2, j)$  if and only if  $x_1 x_2 \in E(G)$  and  $(x_1, i) < (x_2, j)$ . We denote the unwinding of  $G$  with respect to the ordering  $<$  by  $u_{<}(G)$ .

<sup>19</sup>Note that Cook's [2] proof of his unwinding theorem regarding  $F$ -systems works for unwindings of infinitary systems as well as finitary ones. The proof strategy used here does not extend to the unwinding of all infinitary systems—the cases where it would fail in the infinite case are related to the discussion of the elimination of self-reference (and the Uniformity Constraint) in [10].

<sup>20</sup>Cook [2] attributes the basic idea of “unwinding” to Thomas Bolander who we know (from personal communication) also has some unpublished results on the graph-theoretic nature of paradox supporting structures. We in turn owe the idea of unwinding to Bolander via [2].



**Fig. 9** The 3-cycle and its unwinding at various stages

**Lemma 7** *Let  $G$  be a finite directed graph and let  $<$  be a total order on  $V(G)$ . Then  $u_{<}(G)$  is acyclic.*

*Proof* Assume there is a directed cycle in  $u_{<}(G)$  with vertices  $(x_1, i_1), (x_2, y_1), \dots, (x_k, i_k), (x_1, i_1)$ . Then by the definition of the edge set of  $u_{<}(G)$ , we have  $(x_1, i_1) < (x_2, y_1) < \dots < (x_k, i_k) < (x_1, i_1)$ , a contradiction. Hence there is no directed cycle. □

**Lemma 8** *Let  $G$  be a finite directed graph and let  $<$  be a total order on  $V(G)$ . Then  $G$  is dangerous (precarious) if and only if  $u_{<}(G)$  is dangerous (precarious).*

*Proof* Let  $d$  be a denotation assignment on  $V(G)$ . Define a denotation assignment  $d'$  on  $V(G) \times \mathbb{N}$  as follows. For  $x \in V(G)$  and  $k \in \mathbb{N}$ , put

$$T_{x,k} = d(x) [t \Rightarrow (t, k + 1) \mid t \leq x] [t \Rightarrow (t, k) \mid t > x].$$

That is, in the sentence  $d(x)$ , we replace each  $t \in V(G)$  which is at most  $x$  with  $(t, k + 1) \in V(G) \times \mathbb{N}$  and then replace each  $t \in V(G)$  which is greater than  $x$  with  $(t, k) \in V(G) \times \mathbb{N}$ . This gets the referencing to match the unwinding graph.

Now for  $(x, k) \in V(G) \times \mathbb{N}$ , let  $d'((x, k)) = \bigwedge_{j \geq k} T_{x,j}$ . Then, by construction,  $\mathcal{G}_{V(G) \times \mathbb{N}, d'} = u_{<}(G)$ .

We first show that if  $G$  is not dangerous (precarious) then  $u_{<}(G)$  is not dangerous (precarious). Assume  $G$  is not dangerous (precarious). Let  $v$  be

an acceptable truth assignment on  $V(G)$  with respect to  $d$ . Define a truth assignment  $v'$  on  $V(G) \times \mathbb{N}$  by setting  $v'((x, k)) = v(x)$ . Then

$$\begin{aligned} \llbracket d'((x, k)) \rrbracket(v') &= \bigwedge_{j \geq k} \llbracket T_{x,j} \rrbracket(v') \\ &= \bigwedge_{j \geq k} d(x) [t \Rightarrow v'((t, k + 1)) \mid t \leq x] [t \Rightarrow v'((t, k)) \mid t > x] \\ &= \bigwedge_{j \geq k} d(x) [t \Rightarrow v(t) \mid t \leq x] [t \Rightarrow v(t) \mid t > x] \\ &= \bigwedge_{j \geq k} \llbracket d(x) \rrbracket(v) \\ &= \llbracket d(x) \rrbracket(v) \\ &= v(x) \\ &= v'((x, k)) \end{aligned}$$

Hence  $v'$  is acceptable with respect to  $d'$ . Hence  $u_{<}(G)$  is not dangerous. To see that  $u_{<}(G)$  is not precarious, note that if  $v_1$  and  $v_2$  are distinct acceptable truth assignments on  $V(G)$ , then  $v'_1$  and  $v'_2$  are distinct as well.

Now for the other direction, assume that  $u_{<}(G)$  is not dangerous (precarious). Let  $v'$  be an acceptable truth assignment on  $V(G) \times \mathbb{N}$  with respect to  $d'$ .

We claim that for each  $x \in V(G)$ , we have  $v'((x, k_1)) = v'((x, k_2))$  for all  $k_1, k_2 \in \mathbb{N}$ . Let  $x \in V(G)$ . Assume  $v'((x, k)) = 1$  for some  $k \in \mathbb{N}$  and let  $j \geq k$ . We have  $1 = v'((x, k)) = \llbracket d'((x, k)) \rrbracket(v') = \llbracket d'(x, j) \wedge \bigwedge_{k \leq i < j} T_{x,i} \rrbracket(v')$ . Hence  $v'(x, j) = \llbracket d'(x, j) \rrbracket(v') = 1$ . Hence for each  $x \in V(G)$  we have  $M_x \in \mathbb{N}$  such that either  $v'((x, j)) = 1$  for all  $j \geq M_x$  or  $v'((x, j)) = 0$  for all  $j \in \mathbb{N}$ . Since  $G$  is finite, we can define  $M = \max_{x \in V(G)} M_x$ . To get a contradiction, assume there is some  $y \in V(G)$  and  $k \in \mathbb{N}$  such that  $v'((y, k)) \neq v'((y, M))$ . Then from the above we know that  $k < M$ , so we may let  $(z, j)$  be the maximum such pair. But by the definition of  $d'$ , the constituents of  $d'((z, j))$  are only pairs  $(x, i)$  with  $(x, i) > (z, j)$ . By the maximality of  $(z, j)$  we have  $v'((x, i)) = v'((x, M))$  for any such  $(x, i)$ . Hence  $v'((z, j)) = \llbracket d'((z, j)) \rrbracket(v') = \llbracket d'((z, M)) \rrbracket(v') = v'((z, M))$  which is the desired contradiction. This proves the claim.

Define a truth assignment  $v$  on  $V(G)$  by letting  $v(x) = v'((x, 0))$  for each  $x \in V(G)$ . This assignment is acceptable with respect to  $d$  since

$$\begin{aligned} v(x) &= v'((x, 0)) \\ &= \llbracket d'((x, 0)) \rrbracket(v') \\ &= \bigwedge_{j \geq 0} \llbracket d(x) [t \Rightarrow (t, j + 1) \mid t \leq x] [t \Rightarrow (t, j) \mid t > x] \rrbracket(v') \end{aligned}$$

$$\begin{aligned}
 &= \bigwedge_{j \geq 0} d(x)[t \Rightarrow v'((t, j + 1)) \mid t \leq x][t \Rightarrow v'((t, j)) \mid t > x] \\
 &= \bigwedge_{j \geq 0} d(x)[t \Rightarrow v'((t, j)) \mid t \leq x][t \Rightarrow v'((t, j)) \mid t > x] \\
 &= \bigwedge_{j \geq 0} d(x)[t \Rightarrow v'((t, 0)) \mid t \in V(G)] \\
 &= d(x)[t \Rightarrow v'((t, 0)) \mid t \in V(G)] \\
 &= \llbracket d(x) \rrbracket (v).
 \end{aligned}$$

Thus  $G$  is not dangerous. By our claim above, different acceptable truth assignments on  $V(G) \times \mathbb{N}$  will give different acceptable truth assignments on  $V(G)$ . Hence  $G$  is not precarious either.  $\square$

### 4 Groundedness

Herzberger [6] introduced a notion of “groundlessness” which was meant to capture the way in which some sentences “suffer from *unconsummated reference* much like the bureaucratic regress in which each clerk endlessly refers you to the next clerk to settle your accounts”.<sup>21</sup> Herzberger’s idea was that each sentence has a “domain”, where this is understood as the set of things it is *about*. And some sentences have domains that include sentences, whose domain include sentences, etc. If this chain of aboutness never ends then the sentence is groundless.

The relation between a sentence  $S$  and its domain  $D(S)$  is sensitive to some of the same factors that are operative in general set theory. In case some members of  $D(S)$  themselves are sentences, they in turn will have their own domains, which collectively can be designated  $D^2(S)$ : the aggregate of the domains of all sentences in the domain of  $S$ . And it can happen that some members of  $D^2(S)$  are sentences, and so on. Any sentence for which this process fails to terminate will be called “groundless”: ‘ $S$  is groundless’ abbreviates ‘for each integer  $k$ ,  $D^k(S)$  is nonempty’. [6]

Herzberger relies on the intuitions of “aboutness” to give content to  $D$ . With our technology we can make the idea of a sentence domain precise. For a given sentence system  $(S, d)$  we can understand the domain of a sentence

---

<sup>21</sup>Herzberger [6, p. 150]. Of course, [7] also makes use of and rigorously defines a notion of “groundedness” but Kripke’s definition is tied up with his definitions of *jumps* and *fixed-points*, so its much easier to see that Herzberger’s definition is an ancestor of our definition here.

name  $\alpha \in \mathbf{S}$ ,  $D(\alpha) = N^+(\alpha)$  (i.e. the set of  $\alpha$ 's out-neighbors). In general we can define  $D^k(\alpha)$  for each  $k$  as follows:<sup>22</sup>

$$D^0(\alpha) = \{\alpha\},$$

$$\text{For } k \geq 1, D^k(\alpha) = N^+(D^{k-1}(\alpha)).$$

Metaphorically, if a sentence is “grounded” then that excludes the ability to start at it and walk along references forever. And this is clearly the notion that Herzberger is trying to capture with “groundlessness”.

**Definition 18** For a system  $(\mathbf{S}, d)$  and  $\alpha \in \mathbf{S}$ ,  $\alpha$  is *Herzberger-groundless* if and only if for each  $k$ ,  $D^k(\alpha)$  is non-empty.

His definition, however, only captures this idea if sentences are restricted to finite length. When we are dealing with sentences of infinite length, then there are sentences, which are Herzberger-groundless, but which don't involve cycles or infinite paths. To see this consider the sentences system  $(\mathbf{S}, d)$ , where  $\mathbf{S} = \{Z, X_1^1, X_1^2, X_2^2, X_2^3, X_3^3, \dots\}$  and  $d(Z) = \bigwedge_{i \geq 1} \bigwedge_{1 \leq j \leq i} X_j^i$  and for each  $X_j^i$  with  $1 \leq j < i$ ,  $d(X_j^i) = X_{j+1}^i$ . Here  $Z$  is Herzberger-groundless since for every natural number  $k$  there is a chain of references of length  $k$  to the sentence name  $X_k^k$ , so for any  $k$ ,  $D^k(\alpha)$  is non-empty. But there are no cycles and every path emanating from  $Z$  terminates at some  $k$  (see Fig. 10).

A more useful and appropriate definition of *ungroundedness*, however, is close at hand. Let's first define a *ray*, which is an infinite set of vertices connected by an infinite chain of directed edges.

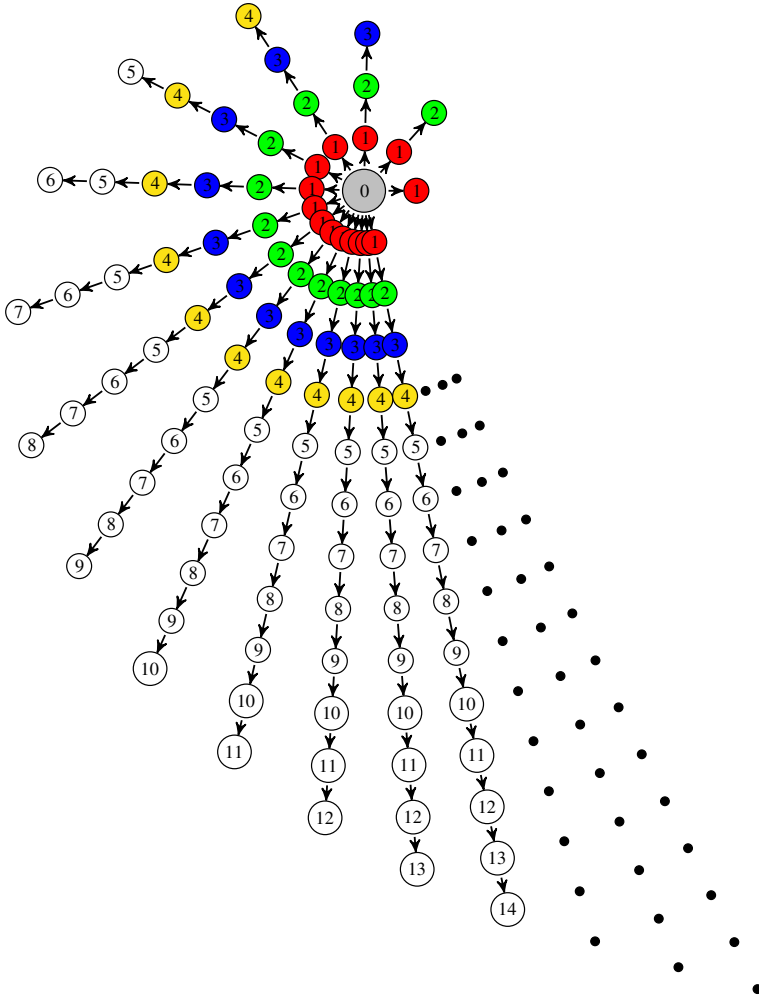
**Definition 19** A *ray* in a directed graph  $G$  is a subgraph with vertex set  $\{v_i\}_{i < \omega}$  and edge set  $\{(v_i, v_{i+1})\}_{i < \omega}$ .

**Definition 20** A directed graph  $G$  is *ungrounded* if it contains a ray or a directed cycle. Otherwise  $G$  is *grounded*.<sup>23</sup>

Intuitively, if  $G$  is grounded, then we can obtain an acceptable truth assignment for any denotation assignment on  $V(G)$  by repeatedly substituting the values of constant sentences in for their names. Since we don't make any arbitrary choices in this process, the constructed acceptable truth assignment should be unique. To make performing this operation infinitely many times precise we will apply Zorn's lemma. Later we will use a similar proof idea to prove a more general result.

<sup>22</sup>Or as:  $D^k(\alpha) = N^+(N^+(N^+ \dots N^+(\alpha)))$ , where the  $N^+$  is iterated  $k$  times.

<sup>23</sup>An anonymous referee suggests the following equivalent definition of ungrounded: a directed graph  $G$  is *ungrounded* if  $G$  has a subgraph  $H$  in which every vertex has positive out-degree.



**Fig. 10** The black hole graph

**Zorn’s Lemma** *Every partially ordered set, in which every chain (i.e. totally ordered subset) has an upper bound, contains at least one maximal element.*

**Definition 21** Let  $G$  be a directed graph. We say that  $A \subseteq V(G)$  is *self-contained* in  $G$  if there are no edges in  $G$  directed from  $A$  to  $G - A$ .

**Lemma 9** *If a directed graph  $G$  is grounded, then it is not dangerous and not precarious.*

*Proof* Assume  $G$  is grounded. Let  $d$  be a denotation assignment on  $V(G)$  such that  $G = \mathcal{G}_{V(G),d}$ . We will show that there is a unique acceptable truth

assignment on  $V(G)$  with respect to  $d$ . Since  $d$  was arbitrary, it follows that  $G$  is not dangerous and not precarious.

For  $A \subseteq V(G)$  which is self-contained in  $G$ , let  $d_A$  be  $d$  restricted to  $A$ . Then  $d_A$  is a denotation assignment on  $A$ . If  $A$  has a unique acceptable truth assignment with respect to  $d_A$ , then we call this truth assignment  $v_A$  and call the pair  $(A, v_A)$  *solved*.

Let  $X$  be the collection of all solved pairs. Define a partial order  $<$  on  $X$  by  $(A, v_A) < (B, v_B)$  if and only if  $A \subsetneq B$  and  $v_A$  is  $v_B$  restricted to  $A$ .

To apply Zorn’s lemma to  $(X, <)$ , we need to show that  $X \neq \emptyset$  and that every chain in  $(X, <)$  has an upper bound. Since  $(\emptyset, v_\emptyset) \in X$  we see that  $X \neq \emptyset$ . Now let  $(A_1, v_{A_1}) < (A_2, v_{A_2}) < \dots$  be an arbitrary chain in  $(X, <)$ . Put  $U = \bigcup_{i>0} A_i$ . Plainly,  $U$  is self-contained in  $G$ . For  $u \in U$ , let  $h(u)$  be the smallest  $i > 0$  such that  $u \in A_i$ . Now, for  $u \in U$ , let  $v_U(u) = v_{A_{h(u)}}(u)$ . We claim that  $(U, v_U)$  is an upper bound for the chain. By definition  $A_i \subseteq U$  and  $v_{A_i}$  is  $v_U$  restricted to  $A_i$  for each  $i > 0$ . It remains to be shown that  $v_U$  is the unique acceptable truth assignment on  $U$  with respect to  $d_U$ . Assume  $v_U$  is not acceptable and pick  $u \in U$  with  $h(u)$  minimal such that  $v_U(u) \neq \llbracket d_U(u) \rrbracket(v_U)$ . Put  $B = A_{h(u)}$ . Then

$$v_B(u) = v_U(u) \neq \llbracket d_U(u) \rrbracket(v_U) = \llbracket d_B(u) \rrbracket(v_B) = v_B(u).$$

This is a contradiction. Hence  $v_U$  is acceptable. To see that  $v_U$  is unique, assume there is a different acceptable truth assignment on  $U$  with respect to  $d_U$ , call it  $v_O$ . Take  $u \in U$  with  $h(u)$  minimal such that  $v_U(u) \neq v_O(u)$ . Again put  $B = A_{h(u)}$ . Then  $v_O$  restricted to  $B$  is an acceptable truth assignment on  $B$  with respect to  $d_B$  which is different from  $v_B$ . This is a contradiction. Thus we conclude that  $(U, v_U)$  is an upper bound for the chain.

Applying Zorn’s lemma gives us a solved pair  $(M, v_M)$  which is maximal in  $(X, <)$ . We will show that  $M = V(G)$  and hence  $v_M$  is the desired unique acceptable truth assignment on  $V(G)$  with respect to  $d$ . So assume  $M \neq V(G)$ . Put  $J = V(G) - M$ .

First assume that there is some  $z \in J$  such that  $T = M \cup \{z\}$  is self-contained. Since  $T$  is self-contained,  $d(z)$  involves only elements of  $M$ . Thus, letting  $v_T(x) = v_M(x)$  for each  $x \in M$  and  $v_T(z) = d(z)[x \Rightarrow v_M(x) \mid x \in M]$  makes  $v_T$  the unique acceptable truth assignment on  $T$  with respect to  $d_T$ . We conclude that  $(T, v_T) \in X$  and  $(M, v_M) < (T, v_T)$  contradicting the maximality of  $(M, v_M)$ .

Thus we may assume that  $N^+(z) \cap J \neq \emptyset$  for each  $z \in J$ . Pick  $z_0 \in J$ . For  $k > 0$ , let  $z_k \in N^+(z_{k-1}) \cap J$ . Since  $G$  is acyclic,  $z_0 z_1 z_2 \dots$  is a ray in  $G$  contradicting the fact that  $G$  is grounded. Whence  $M = V(G)$  and the proof is complete.  $\square$

**Lemma 10** *If a directed graph  $G$  contains a directed cycle, then it is both dangerous and precarious.*

*Proof* By Lemma 2, we can assume that  $G$  is a directed cycle. Let  $V(G) = \{v_1, \dots, v_k\}$ . Let  $d$  be a denotation assignment on  $V(G)$  with  $d(v_i) = \neg v_{i+1}$  for



$i < k$ ,  $d(v_k) = \neg v_1$  if  $k$  is odd and  $d(v_k) = v_1$  if  $k$  is even. Then  $G = \mathcal{G}_{V(G),d}$  and  $(V(G), d)$  is paradoxical. Hence  $G$  is dangerous. To see that  $G$  is precarious, just consider the denotation assignment  $d'$  on  $V(G)$  with  $d'(v_i) = v_{i+1}$  for  $i < k$  and  $d(v_k) = v_1$ . Both the truth assignment setting all the  $v_i$  to 1 and the truth assignment setting all the  $v_i$  to 0 are acceptable.  $\square$

**Theorem 11** *A directed graph  $G$  is precarious if and only if it is ungrounded.*

*Proof* Lemma 9 gives the forward implication. For the reverse implication, by Lemmas 2 and 10 we only need to consider the case of  $G$  being a ray with vertex set  $\{v_i\}_{i < \omega}$ . Let  $d$  be a denotation assignment on  $V(G)$  with  $d(v_i) = v_{i+1}$ . Then  $G = \mathcal{G}_{V(G),d}$  and  $(V(G), d)$  is hypodoxical because we get two acceptable truth-value assignments by setting all the  $v_i$  to 1 or setting all the  $v_i$  to 0.  $\square$

**Corollary 12** *If a directed graph is dangerous then it is precarious.*

In the finite case, we get the following complete characterization of danger.

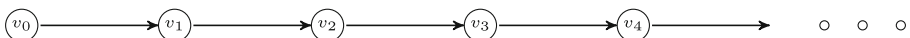
**Corollary 13** *Let  $G$  be a finite directed graph. The following are equivalent:*

- (a)  $G$  is dangerous;
- (b)  $G$  is precarious;
- (c)  $G$  contains a directed cycle;
- (d)  $G$  contains a subdivision of the Liar graph.

The results above exhaust the relations between ungroundedness, cyclicity, precarity and danger; that is, there are ungrounded graphs that are not dangerous (the ray in Fig. 11), precarious graphs that are not dangerous (again the ray in Fig. 11) and dangerous graphs that contain no cycle (the Yablo graph in Fig. 3).

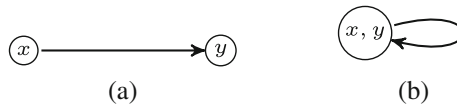
**Definition 22** *A homomorphism from a directed graph  $G$  to a directed graph  $H$  is a function  $f : V(G) \rightarrow V(H)$  such that if  $xy \in E(G)$ , then  $f(x)f(y) \in E(H)$ . If  $f$  is a bijection, and  $f^{-1} : V(H) \rightarrow V(G)$  is a homomorphism, then we call  $f$  an isomorphism. Additionally, we call the graph with vertex set  $f(V(G))$  and edge set  $\{f(x)f(y) \mid xy \in E(G)\}$  the homomorphic image of  $G$  under  $f$ .*

In Fig. 12 we see an example of a non-isomorphic homomorphic image. Additionally, since a path of length one is neither precarious nor dangerous, this example shows that taking of a homomorphic image can introduce precarity



**Fig. 11** A ray

**Fig. 12** The Liar graph is a homomorphic image of the path of length one



and danger. However, precariousity cannot be lost by taking a homomorphic image as the next lemma shows.

**Lemma 14** *Let  $G$  be a directed graph. Then  $G$  is precarious if and only if every homomorphic image of  $G$  is precarious.*

*Proof* Since  $G$  is a homomorphic image of itself, the reverse implication is trivial. To prove the forward implication, assume  $G$  is precarious. Then, by Theorem 11  $G$  contains a directed cycle or a ray. Let  $H$  be a homomorphic image of  $G$  and let  $f : V(G) \rightarrow V(H)$  be a homomorphism. If there is a directed path between  $x, y \in V(G)$ , then we cannot have  $f(x) = f(y)$ , for otherwise  $H$  would contain a directed cycle and hence be precarious. In particular,  $G$  cannot contain a directed cycle. Thus  $G$  contains a ray  $\{v_i\}_{i < \omega}$ . If  $f(v_i) = f(v_j)$  for  $i \neq j$ , then  $H$  contains a directed cycle and we are done. Otherwise  $H$  contains a ray and we are done.  $\square$

**Conjecture 15** *Let  $G$  be a directed graph. Then  $G$  is dangerous if and only if every homomorphic image of  $G$  is dangerous.*

### 5 Compactness and Dangerous Tails

#### 5.1 Locally Finite Graphs

Using the compactness theorem of first-order logic, we will show that graphs for which  $N^+(v)$  is finite for every vertex  $v$  are not dangerous.

**Gödel Compactness** *A set of first-order sentences has a model if and only if every finite subset of it has a model.*

To apply this theorem we need to be careful since a first-order sentence must have finite length. The following generalization of Lemma 1 gives us the control over the lengths of sentences that we need.

**Definition 23** For  $v \in V_S$  and  $I \subseteq S$ , let

$$v^I = \{u \in V_S \mid \forall \alpha \in S - I u(\alpha) = v(\alpha)\}.$$

**Definition 24** Let  $g$  be a function from  $V_S$  to  $\{0, 1\}$ . We say that  $g$  is independent of  $I \subseteq S$  if  $g$  is constant on  $v^I$  for each  $v \in V_S$ .

Note that every such function is independent of the empty set.

**Lemma 16** *For any function  $g$  from  $V_S$  to  $\{0, 1\}$  and any  $I \subseteq S$  of which  $g$  is independent, there exists a sentence  $\zeta_{g,I} \in S^+$  with the following properties:*

- no sentence name from  $I$  appears in  $\zeta_{g,I}$ ;
- $\zeta_{g,I} = g$ ;
- $\zeta_{g,I}$  has finite length if  $S - I$  is finite.

*Proof* Let  $g$  be a function from  $V_S$  to  $\{0, 1\}$ . First, if  $g$  is a constant function, put  $\zeta_{g,I} = \top$  if  $g$  maps everything to 1 and  $\zeta_{g,I} = \perp$  if  $g$  maps everything to 0.

Otherwise  $S - I$  has at least one element and we proceed as follows. Let  $B = \{v \in V_S \mid \forall \alpha \in I v(\alpha) = 0\}$ . Note that  $g$  is completely determined by its values on the elements of  $B$ .

For  $v \in V_S$  and  $\alpha \in S$ , let

$$P(v, \alpha) = \begin{cases} \alpha & \text{if } v(\alpha) = 1 \\ \neg\alpha & \text{if } v(\alpha) = 0 \end{cases}.$$

Let  $C = \{v \in B \mid g(v) = 1\}$ . For  $v \in C$ , define

$$\chi_v = \bigwedge_{\alpha \in S-I} P(v, \alpha).$$

Note that  $\chi_v(r) = g(r)$  if  $r = v$  and  $\chi_v(r) = 0$  if  $r \neq v$ . Let  $\zeta_{g,I}$  be the sentence

$$\bigvee_{v \in C} \chi_v.$$

Then, for any  $r \in V_S$ , we have

$$\llbracket \zeta_{g,I} \rrbracket(r) = \left[ \bigvee_{v \in C} \chi_v \right] (r) = \bigvee_{v \in C} \llbracket \chi_v \rrbracket(r) = \llbracket \chi_r \rrbracket(r) = g(r).$$

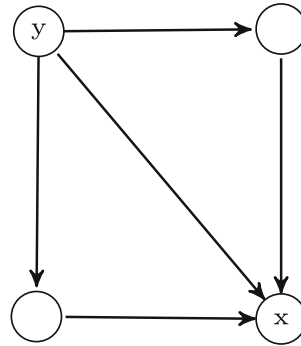
Hence  $\zeta_{g,I} = g$ . Now the length of  $\zeta_{g,I}$  is at most  $|C||S - I| \leq |B||S - I| \leq 2^{|\mathbf{S}-I|}|\mathbf{S} - I|$ . Thus  $\zeta_{g,I}$  has finite length if  $S - I$  is finite. □

**Lemma 17** *Let  $G$  be a directed graph such that  $|N_G^+(v)|$  is finite for every  $v \in V(G)$ . If  $G$  is dangerous then some finite subgraph of  $G$  is dangerous.*

*Proof* Assume that no finite subgraph of  $G$  is dangerous and let  $d$  be a denotation assignment on  $V(G)$  such that  $G = \mathcal{G}_{V(G),d}$ . For each  $v \in V(G)$ , put  $I_v = V(G) - N_G^+(v)$ . Construct a first order language  $\mathcal{L}$  as follows. The constants of  $\mathcal{L}$  are the vertices of  $G$  together with  $\top$  and  $\perp$ . The axioms are the following.

- $\top \neq \perp$
- $x = \top \vee x = \perp$  for each  $x \in V(G)$ ,
- $x = \zeta_{\llbracket d(x) \rrbracket, I_x}$  for each  $x \in V(G)$ .

**Fig. 13** A graph with sink  $x$  and source  $y$



By Lemma 16 each of the sentences is of finite length. Note that a model of the language is an acceptable truth assignment on  $V(G)$  relative to  $d$ . Since no finite subgraph of  $G$  is dangerous, every finite subset of the axioms has a model. Thus, by compactness, the whole language has a model and hence  $(V(G), d)$  is not paradoxical. Since  $d$  was arbitrary, we conclude that  $G$  is not dangerous.<sup>24</sup>  $\square$

**Corollary 18** *Any directed acyclic graph  $G$  such that  $|N_G^+(v)|$  is finite for every  $v \in V(G)$  is not dangerous.*

*Proof* Combine Corollary 13 and Lemma 17.  $\square$

## 5.2 Topological Sorting

Now we show that if  $G$  is an acyclic directed graph then its vertices can be ordered left to right such that edges only go to the right.

**Definition 25** Let  $G$  be a directed graph. A vertex  $v \in V(G)$  is called a *sink* if  $N_G^+(v)$  is empty and a *source* if  $N_G^-(v)$  is empty (Fig. 13).

**Definition 26** Let  $G$  be a directed graph. A *topological sort* on  $G$  is a total ordering  $<$  of  $V(G)$  such that if  $(a, b) \in E(G)$ , then  $a < b$ .

We prove the following easy lemma for completeness.

**Lemma 19** *If  $G$  is a finite directed acyclic graph, then  $G$  has a topological sort.*

<sup>24</sup>The proof technique used here is very similar to the compactness-based proof of the De Bruijn-Erdős coloring theorem [1] stating that an infinite graph can be  $k$ -colored if and only if each of its finite subgraphs can be  $k$ -colored. Lemma 17 can also be proved (as can the De Bruijn-Erdős theorem) by directly applying Zorn's lemma as we need to do in the more complicated results below.

*Proof* Assume (to reach a contradiction) that the lemma is false and let  $G$  be a counterexample with the minimum number of vertices. Since  $G$  is finite it has a source  $v$ . By minimality  $G - v$  has a topological sort  $\{v_1, \dots, v_r\}$ . But then  $\{v, v_1, \dots, v_r\}$  is a topological sort of  $G$ . This is a contradiction.  $\square$

**Lemma 20** *Let  $G$  be an directed acyclic graph. Then  $G$  has a topological sort.*

*Proof* We construct a first order language  $\mathcal{L}$  and apply the compactness theorem. Let the elements of  $V(G)$  be the constants of  $\mathcal{L}$  and let  $R$  be  $\mathcal{L}$ 's only relation symbol. Define the axioms of  $\mathcal{L}$  as follows.

- $\neg(a = b)$  for all distinct  $a, b \in V(G)$ ,
- $aRb$  for all  $(a, b) \in E(G)$ ,
- $aRb \rightarrow \neg(bRa)$  for all  $a, b \in V(G)$ ,
- $aRb \vee bRa \vee a = b$  for all  $a, b \in V(G)$ ,
- $(aRb \wedge bRc) \rightarrow aRc$  for all  $a, b, c \in V(G)$ .

Let  $A$  be a finite subset of the axioms. Let  $C$  be the set of all constants appearing in some axiom of  $A$ . Create  $A'$  from  $A$  by adding in all axioms involving only the elements of  $C$ . Then  $A'$  is still finite and if  $A'$  has a model, so does  $A$ . Put  $H = G[C]$ . Since  $H$  is finite and acyclic it has a topological sort  $<$ . Letting  $<$  be the interpretation of  $R$  gives a model of  $A'$  and hence  $A$ .

Thus, by the compactness theorem, the entire set of axioms has a model. The interpretation of  $R$  in this model is the desired topological sort.  $\square$

### 5.3 Dangerous Tails

**Lemma 21** *Let  $G$  be a directed graph. If for every induced subgraph  $H$  of  $G$  there exists  $\emptyset \neq C \subseteq V(H)$  which is self-contained in  $H$  such that  $G[C]$  is not dangerous, then  $G$  is not dangerous.*

*Proof* Assume that for every induced subgraph  $H$  of  $G$  there exists  $\emptyset \neq C \subseteq V(H)$  which is self-contained in  $H$  such that  $G[C]$  is not dangerous. Let  $d$  be a denotation assignment on  $V(G)$  such that  $G = \mathcal{G}_{V(G),d}$ . We will show that there is an acceptable truth assignment on  $V(G)$  with respect to  $d$ . Since  $d$  was arbitrary, it follows that  $G$  is not dangerous.

For  $A \subseteq V(G)$  which is self-contained in  $G$ , let  $d_A$  be  $d$  restricted to  $A$ . Then  $d_A$  is a denotation assignment on  $A$ . If  $A$  has an acceptable truth assignment with respect to  $d_A$ , then we pick an acceptable truth assignment  $v_A$  and call the pair  $(A, v_A)$  solved.

Let  $X$  be the collection of all solved pairs. Define a partial order  $<$  on  $X$  by  $(A, v_A) < (B, v_B)$  if and only if  $A \subsetneq B$  and  $v_A$  is  $v_B$  restricted to  $A$ .

To apply Zorn's lemma to  $(X, <)$ , we need to show that  $X \neq \emptyset$  and that every chain in  $(X, <)$  has an upper bound. Since  $(\emptyset, v_\emptyset) \in X$  we see that  $X \neq \emptyset$ . Now let  $(A_1, v_{A_1}) < (A_2, v_{A_2}) < \dots$  be an arbitrary chain in  $(X, <)$ . Put  $U = \bigcup_{i>0} A_i$ . Plainly,  $U$  is self-contained in  $G$ . For  $u \in U$ , let  $h(u)$  be the smallest  $i > 0$  such that  $u \in A_i$ . Now, for  $u \in U$ , let  $v_U(u) = v_{A_{h(u)}}(u)$ . We claim that

$(U, v_U)$  is an upper bound for the chain. By definition  $A_i \subseteq U$  and  $v_{A_i}$  is  $v_U$  restricted to  $A_i$  for each  $i > 0$ . It remains to be shown that  $v_U$  is an acceptable truth assignment on  $U$  with respect to  $d_U$ . Assume  $v_U$  is not acceptable and pick  $u \in U$  with  $h(u)$  minimal such that  $v_U(u) \neq \llbracket d_U(u) \rrbracket(v_U)$ . Put  $B = A_{h(u)}$ . Then

$$v_B(u) = v_U(u) \neq \llbracket d_U(u) \rrbracket(v_U) = \llbracket d_B(u) \rrbracket(v_B) = v_B(u).$$

This is a contradiction. Hence  $v_U$  is acceptable. Thus we conclude that  $(U, v_U)$  is an upper bound for the chain.

Applying Zorn's lemma gives us a solved pair  $(M, v_M)$  which is maximal in  $(X, <)$ . We will show that  $M = V(G)$  and hence  $v_M$  is the desired acceptable truth assignment on  $V(G)$  with respect to  $d$ . So assume  $M \neq V(G)$ . Put  $H = G - M$ . By assumption, we have  $\emptyset \neq C \subseteq V(H)$  which is self-contained in  $H$  such that  $G[C]$  is not dangerous. Put  $B = M \cup C$ . Note that  $B$  is self-contained. Since  $C$  is not dangerous, we can extend  $v_M$  to an acceptable truth assignment  $v_B$  on  $B$  with respect to  $d_B$ . But then  $(B, v_B) \in X$  and  $(B, v_B) > (M, v_M)$  contradicting the maximality of  $(M, v_M)$ . Hence  $M = V(G)$  and the proof is complete.  $\square$

A good way to get self-contained sets in an acyclic graph is to topological sort the graph and take all vertices "to the right" of a given vertex. To make this precise we introduce the concept of a tail.

**Definition 27** Let  $G$  be a directed acyclic graph and let  $<$  be a topological sort on  $G$ . For  $z \in V(G)$ , the  $z$ -tail of  $G$  (with respect to  $<$ ) is the subgraph induced on  $\{x \in V(G) \mid x > z\}$ . An induced subgraph of  $G$  that is a  $z$ -tail for some  $z \in V(G)$  is called a *tail* of  $G$ .

**Lemma 22** Let  $G$  be a directed acyclic graph and let  $<$  be a topological sort on  $G$ . If every induced subgraph of  $G$  has a tail which is not dangerous, then  $G$  is not dangerous.

*Proof* Let  $H$  be an arbitrary induced subgraph of  $G$ . We need to show that there exists  $\emptyset \neq A \subseteq V(H)$  which is self-contained in  $H$  such that  $G[A]$  is not dangerous. Pick  $z \in V(H)$  such that the  $z$ -tail of  $H$  is not dangerous. Let  $T_z$  be the  $z$ -tail of  $H$ . Note that  $T_z$  is self-contained. Thus, if  $T_z$  is non-empty, then we are done. Hence we may assume that  $T_z$  is empty. But then  $x \leq z$  for every  $x \in V(H)$ . Hence  $z$  is a sink in  $H$  and in particular,  $\{z\}$  is a non-empty, self-contained set which induces a non-dangerous graph. This completes the proof.  $\square$

**Corollary 23** If  $G$  is a dangerous directed acyclic graph, then  $|N_G^+(v)|$  is infinite for infinitely many  $v \in V(G)$ .

*Proof* Let  $G$  be a directed acyclic graph  $G$  such that  $|N_G^+(v)|$  is infinite for only finitely many  $v \in V(G)$ . By Lemma 20  $G$  has a topological sort  $<$ . Let  $H$  be an

arbitrary induced subgraph of  $G$ . Since there are only finitely many  $v \in V(H)$  with  $|N_G^+(v)|$  infinite, there is a largest (under the order  $<$ ) such vertex  $z_H$ . Then the  $z_H$ -tail of  $H$  has no vertices with infinite out degree and hence is not dangerous by Lemma 17. Thus every induced subgraph of  $G$  has a non-dangerous tail. Applying Lemma 22 finishes the proof.  $\square$

### 6 Reciprocity and Underlying Graphs

Let  $S = \{A_1, A_2, \dots\}$  and  $d(A_i) = \neg A_{i+1}$ . Then the reference graph of the sentence system  $(S, d)$  is the ray and we know that this is precarious but not dangerous from above. Intuitively, in the reference graph we have an edge from  $A_1$  to  $A_2$  because  $d(A_1)$  is “a function of”  $A_2$ .<sup>25</sup> In this case, fixing the value of  $A_2$  determines what the value of  $A_1$  must be. But it is also true that fixing the value of  $A_1$  determines what the value of  $A_2$  must be—so  $A_1$  and  $A_2$  are reciprocal. Looking at it this way, we would want to see an edge going in both directions. We can capture these intuitive ideas by considering necessary and sufficient conditions for danger on the underlying *undirected* graph of a directed graph  $G$ .

**Definition 28** Let  $G$  be a directed graph. The *underlying undirected graph* of  $G$  is the graph  $\mathcal{U}(G)$  with vertex set  $V(G)$  and an edge between  $x, y \in V(G)$  for each  $xy \in E(G)$  and  $yx \in E(G)$ . We also call  $G$  an *orientation* of  $\mathcal{U}(G)$ .<sup>26</sup>

It turns out that we can completely classify the undirected graphs which have dangerous orientations—they are precisely the ones containing a cycle.

**Theorem 24** *A graph has a dangerous orientation if and only if it contains a cycle.*

*Proof* The reverse direction is easy, since if  $F$  is a graph that contains a cycle we may orient the edges of the cycle clockwise and the other edges arbitrarily and conclude that the orientation is dangerous using Lemma 10.

For the forward direction, assume  $G$  is a directed graph such that  $\mathcal{U}(G)$  is acyclic. Let  $d$  be a denotation assignment on  $V(G)$  such that  $G = \mathcal{G}_{V(G),d}$ . By Lemma 3 we may assume that there is no “junk”; that is, for every  $x \in V(G)$ , if  $d(x) \notin \{\perp, \top\}$ , then there exist truth assignments  $v_0, v_1$  on  $V(G)$  such that  $\llbracket d(x) \rrbracket(v_0) = 0$  and  $\llbracket d(x) \rrbracket(v_1) = 1$ . Also, without loss of generality, we may assume that  $\mathcal{U}(G)$  is connected.

For  $A \subseteq V(G)$ , call  $x \in A$  *interior* to  $A$  if  $x$  has no edges to  $G - A$ . The set of interior vertices of  $A$  is the *interior* of  $A$  and is denoted  $\mathcal{I}(A)$ . Let the

<sup>25</sup>This intuitive idea of “dependence” can only be taken so far. In Appendix C we show that defining a graph based on “dependence” turns out to be useless in general.

<sup>26</sup>Note that if  $G$  contains a cycle of length two through vertices  $x$  and  $y$ —such as in Jourdain’s paradox—then  $\mathcal{U}(G)$  will have two edges between  $x$  and  $y$ .

exterior vertices of  $A$  be  $\mathcal{E}(A) = A - \mathcal{I}(A)$ . Call  $A \subseteq V(G)$  tame if  $\mathcal{U}(G[A])$  is connected and for each  $x \in \mathcal{E}(A)$  we have  $N^+(x) \cap A = \emptyset$ . Additionally, let  $d_A$  be  $d$  restricted to  $A$ . We need to extend the notion of acceptable truth assignment as follows. A function  $v$  from  $A$  to  $\{0, 1\}$  is called acceptable on  $A$  relative to  $d_A$  if for each  $y \in \mathcal{I}(A)$  we have  $\llbracket d_A(y) \rrbracket(v) = v(y)$ . This is well-defined since  $d_A(y)$  involves only elements of  $A$  by the definition of  $\mathcal{I}(A)$ .

Now, for tame  $A \subseteq V(G)$  we call the pair  $(A, v_A)$  solved if  $v_A$  is an acceptable truth assignment (in the extended sense above) on  $A$  with respect to  $d_A$ . Let  $X$  be the set of all solved pairs in  $G$ . Define a partial ordering on  $X$  by  $(A, v_A) < (B, v_B)$  if and only if  $A \subsetneq B$  and  $v_A$  is  $v_B$  restricted to  $A$ .

To apply Zorn's lemma to  $(X, <)$ , we need to show that  $X \neq \emptyset$  and that every chain in  $(X, <)$  has an upper bound. Since  $(\emptyset, v_\emptyset) \in X$  we see that  $X \neq \emptyset$ . Now let  $(A_1, v_{A_1}) < (A_2, v_{A_2}) < \dots$  be an arbitrary chain in  $(X, <)$ . Put  $U = \bigcup_{i>0} A_i$ . Plainly,  $A_i \subseteq U$  for each  $i \geq 1$ .

Now we show that  $U$  is tame. Let  $a, b \in U$ . Then we have  $i_a$  and  $i_b$  such that  $a \in A_{i_a}$  and  $b \in A_{i_b}$ . Let  $i$  be the maximum of  $i_a$  and  $i_b$ . Then  $a, b \in A_i$ . Since  $A_i$  is tame,  $\mathcal{U}(G[A_i])$  is connected. Hence there is a path between  $a$  and  $b$  in  $\mathcal{U}(G[U])$ . Since  $a$  and  $b$  were arbitrary elements of  $U$ , we conclude that  $\mathcal{U}(G[U])$  is connected. Since an exterior vertex of  $U$  must be exterior in each  $A_i$  that contains it, we see that  $N^+(x) \cap U = \emptyset$  for each  $x \in \mathcal{E}(U)$ . Whence  $U$  is tame.

We claim that if  $u \in \mathcal{I}(U)$ , then  $u \in \mathcal{I}(A_i)$  for some  $i \geq 1$ . Pick  $k$  such that  $u \in A_k$ . If  $N^+(u) \cap (U - A_k)$  is empty then we have  $u \in \mathcal{I}(A_k)$ . Otherwise we may pick  $y \in N^+(u) \cap (U - A_k)$ . Since  $u$  is in the interior of  $U$ ,  $y \in A_j$  for some  $j > k$ . But  $A_j$  is tame and  $u \in A_j$ , so we must have either  $u \in \mathcal{I}(A_j)$  or  $N^+(u) \cap A_j = \emptyset$ . The latter is impossible since  $uy \in G[A_j]$ . Hence  $u \in \mathcal{I}(A_j)$ . This proves the claim.

Now we construct an acceptable truth assignment  $v_U$  on  $U$ . For  $u \in U$ , let  $k_u$  be minimal such that  $u \in A_{k_u}$ , and let  $v_U(u) = v_{A_{k_u}}(u)$ . We claim that  $(U, v_U)$  is a solved pair. By definition  $v_{A_i}$  is  $v_U$  restricted to  $A_i$  for each  $i \geq 1$ . Thus it only remains to show that  $v_U$  is an acceptable truth assignment on  $U$ . Take  $y \in \mathcal{I}(U)$  and let  $r \geq 1$  be minimal such that  $y \in \mathcal{I}(A_r)$ . Then  $\llbracket d_U(y) \rrbracket(v_U) = \llbracket d_{A_r}(y) \rrbracket(v_{A_r}) = v_{A_r}(y) = v_U(y)$  since  $v_{A_r}$  is acceptable on  $A_r$ . Hence  $v_U$  is acceptable on  $U$  and thus  $(U, v_U)$  is a solved pair.

Thus every chain in  $(X, <)$  has an upper bound and Zorn's lemma gives us a maximal element  $(M, v_M) \in X$ . If  $M = V(G)$ , then  $v_M$  is an acceptable truth assignment on  $V(G)$  with respect to  $d$  and we are done. Thus assume that  $M \neq V(G)$ .

First if  $\mathcal{E}(M) \neq \emptyset$  then pick  $z \in \mathcal{E}(M)$ . Put  $M' = M \cup N^+(z)$ . Then  $z \in \mathcal{I}(M')$ . Since  $\mathcal{U}(G[M'])$  is connected and  $\mathcal{U}(G)$  is acyclic we see that for each  $x \in \mathcal{E}(M')$  we have  $N^+(x) \cap M' = \emptyset$ . Additionally, it is clear that  $\mathcal{U}(G[M'])$  is connected. Hence  $M'$  is tame. Since there is no "junk" we can define an acceptable truth assignment  $v'$  on  $M'$  by letting  $v'(x) = v_M(x)$  for  $x \in M$  and choosing the values of  $v'$  on  $N^+(z)$  so that  $\llbracket d(z) \rrbracket(v') = v_M(z)$ . But then  $(M', v') \in X$  and  $(M', v') > (M, v_M)$  contradicting the maximality of  $(M, v_M)$ .



Hence we may assume that  $M = \mathcal{I}(M)$ . If  $M \neq \emptyset$ , then since  $\mathcal{U}(G)$  is connected we have  $z \in V(G) - M$  and  $y \in M$  such that  $zy \in E(G)$ . If  $M = \emptyset$ , then pick  $z \in V(G)$  arbitrarily. Put  $M' = M \cup \{z\} \cup N^+(z)$ . Since  $\mathcal{U}(G[M'])$  is connected and  $\mathcal{U}(G)$  is acyclic we see that for each  $x \in \mathcal{E}(M')$  we have  $N^+(x) \cap M' = \emptyset$ . If  $M = \emptyset$ , then  $\mathcal{U}(G[M'])$  is clearly connected, otherwise since  $zy \in E(G)$  we see that  $\mathcal{U}(G[M'])$  is connected. Hence  $M'$  is tame. Extend  $v_M$  to a truth assignment  $v'$  on  $M'$  by letting  $v'(x) = 0$  for each  $x \in N^+(z) - \{y\}$  and letting  $v'(z)$  be the resulting forced value. But then  $(M', v') \in X$ , and  $(M', v') > (M, v_M)$ , which contradicts the maximality of  $(M, v_M)$ .

Thus  $V(G) = M$ , and  $v_M$  is an acceptable truth assignment on  $V(G)$ . Hence  $G$  is not dangerous. □

The philosophical implications of this theorem remain to be determined. But it seems that there is some sense in which cyclic structure is required for paradoxicality.

### Appendix A: The Global Function

We briefly mention an equivalent formulation of a paradoxical (hypodoxical) pair in terms of fixed points of functions.

**Definition 29** Let  $\mathbf{S}$  be a set of sentence names. Any function  $f : \mathcal{V}_{\mathbf{S}} \rightarrow \mathcal{V}_{\mathbf{S}}$  gives rise to a denotation assignment  $d_f$  on  $\mathbf{S}$  as follows. For each  $\alpha \in \mathbf{S}$ , let  $f_\alpha : \mathcal{V}_{\mathbf{S}} \rightarrow \{0, 1\}$  be given by  $f_\alpha(v) = f(v)(\alpha)$ . Then put  $d_f(\alpha) = \zeta_{f_\alpha}$  for each  $\alpha \in \mathbf{S}$ .

Going the other direction, for a denotation assignment  $d$  on  $\mathbf{S}$  the *global function*  $\mathcal{F}_{\mathbf{S},d} : \mathcal{V}_{\mathbf{S}} \rightarrow \mathcal{V}_{\mathbf{S}}$  is given by  $\mathcal{F}_{\mathbf{S},d}(v)(\alpha) = \llbracket d(\alpha) \rrbracket(v)$ . Note that these constructions are inverses of each other; that is,  $\mathcal{F}_{\mathbf{S},d_f} = f$  and  $d_{\mathcal{F}_{\mathbf{S},d}} = d$ .

**Lemma 25** *Let  $\mathbf{S}$  be a set of sentence names and  $d$  a denotation assignment on  $\mathbf{S}$ . The pair  $(\mathbf{S}, d)$  is paradoxical (hypodoxical) if and only if  $\mathcal{F}_{\mathbf{S},d}$  has zero (more than one) fixed point.*

*Proof* Just note that  $v \in \mathcal{V}_{\mathbf{S}}$  is a fixed point of  $\mathcal{F}_{\mathbf{S},d}$  if and only if  $v(\alpha) = \mathcal{F}_{\mathbf{S},d}(v)(\alpha) = \llbracket d(\alpha) \rrbracket(v)$  if and only if  $v$  is an acceptable truth assignment on  $\mathbf{S}$  with respect to  $d$ . □

This formulation is quite useful for constructing examples. Let  $\mathbf{S}$  be the natural numbers  $\mathbb{N}$ . We can write each point  $x \in [0, 1]$  as a binary decimal  $0.b_1b_2b_3 \dots$ , and so any function  $f$  from the unit interval to itself gives rise to a denotation assignment  $d_f$  on  $\mathbf{S}$ .<sup>27</sup> Moreover, if  $f$  has no fixed points, then  $d_f$  is paradoxical.

---

<sup>27</sup>For definiteness, for  $x \in [0, 1)$  we take the (unique) binary decimal representation with infinitely many zeros and for  $x = 1$  we take  $0.1111\dots$ . Since this mapping is not surjective, there are some functions from truth value assignments to truth value assignments that are not represented as a function from  $[0, 1]$  to  $[0, 1]$ .

**Appendix B: Subdivisions of Yablo?**

As we saw in Corollary 13, a finite directed graph is dangerous if and only if it contains a subgraph homeomorphic to the Liar graph. It is tempting to think that a simple topological characterization might work in the infinite case as well. The obvious candidate to try is the Yablo graph. However, the following example gives a dangerous graph with no subgraph homeomorphic to the Liar graph and no subgraph homeomorphic to the Yablo graph.

Consider the following setup. Let  $S = \{A_1, A_2, A_3, \dots, B_1, B_2, B_3, \dots\}$  and for each  $A_i \in S$ , let  $d(A_i) = B_i$  and for each  $B_i \in S$ , let  $d(B_i) = \bigwedge_{j>i} \neg A_j$ . So each  $A_i$  says that  $B_i$  is true, while each  $B_i$  says all the  $A_j$  are false, for  $j > i$ .

$$\begin{aligned} d(A_1) &= B_1, & d(B_1) &= \neg A_2 \wedge \neg A_3 \wedge \neg A_4 \wedge \dots \\ d(A_2) &= B_2, & d(B_2) &= \neg A_3 \wedge \neg A_4 \wedge \neg A_5 \wedge \dots \\ d(A_3) &= B_3, & d(B_3) &= \neg A_4 \wedge \neg A_5 \wedge \neg A_6 \wedge \dots \\ & \vdots & & \vdots \end{aligned}$$

There is no acceptable truth assignment for  $(S, d)$ , since if  $v$  is an acceptable truth assignment, then

$$v(A_i) = \llbracket B_i \rrbracket(v) = v(B_i)$$

and

$$v(B_i) = \llbracket \bigwedge_{j>i} \neg A_j \rrbracket(v) = \bigwedge_{j>i} \llbracket \neg A_j \rrbracket(v) = \bigwedge_{j>i} \neg \llbracket A_j \rrbracket(v) = \bigwedge_{j>i} \neg v(A_j) = \bigwedge_{j>i} \neg v(B_j).$$

In particular, for each  $i$ ,

$$v(B_i) = \neg v(B_{i+1}) \wedge \bigwedge_{j>i+1} \neg v(B_j) = \neg v(B_{i+1}) \wedge v(B_{i+1}) = 0.$$

Thus,

$$0 = v(B_0) = \bigwedge_{j>0} \neg v(B_j) = \bigwedge_{j>0} \neg 0 = 1.$$

**Appendix C: Dependence and Reference**

Given a denotation assignment  $d$  on a set  $S$ , we might try to define what it means for  $\alpha \in S$  to “depend on”  $\beta \in S$ . In the above we took a purely syntactic route with reference. We note that if  $\alpha$  does not reference  $\beta$ , then surely  $\alpha$  does not “depend on”  $\beta$  in any direct sense. Can we get a semantic notion of dependence that gives rise to a meaningful dependence structure? The following is the natural definition to try.

**Definition 30** Let  $d$  be a denotation assignment on a set  $S$ . We say that  $\alpha \in S$  depends on  $\beta \in S$  if there exist truth assignments  $v_1, v_2$  on  $S$  which differ only on  $\beta$  such that  $\llbracket d(\alpha) \rrbracket(v_1) \neq \llbracket d(\alpha) \rrbracket(v_2)$ . The dependence graph of  $D_{S,d}$  is the graph with vertex set  $S$  and an edge from  $\alpha \in S$  to  $\beta \in S$  if and only if  $\alpha$  depends on  $\beta$ .

It is not difficult to see that the dependence graph and the reference graph coincide for the  $\mathcal{F}$ -systems studied by Cook and Yablo (see Appendix D). Also, the dependence graph is meaningful for finite sentence systems. However, when we move to the infinite, we can get situations where infinitely many values of  $v$  must be changed in order to change  $\llbracket d(\alpha) \rrbracket(v)$  for a given  $\alpha$ . In particular, we can get paradoxes with the Yablo graph as reference graph which have a dependence graph with no edges at all. In these cases, we cannot tell anything useful about the possibility of a paradox by looking at the dependence graph. We give two examples of this phenomenon.

For the first example, consider a countably infinite list of sentences, where each one is true if and only if infinitely many of the sentences after it are false. We can encode this in a sentence system as follows. Let  $S = \{A_1, A_2, \dots\}$  and define a denotation assignment  $d$  on  $S$  by  $d(A_k) = \neg \bigvee_{i>k} \bigwedge_{j \geq i} A_j$ , for each  $k$ . We leave it as an exercise to check that this is indeed a paradox. The dependence graph has no edges because for any truth assignment, toggling the value of a single sentence will not affect the truth-value of any other sentence.

The second example is based on the fact that flipping only finitely many bits in the binary representation of a real number cannot change it from being rational to irrational or vice-versa. Let  $[0, 1] \subseteq \mathbb{R}$  denote the unit interval. The function  $f : [0, 1] \rightarrow [0, 1]$  given by  $f(x) = 0$  if  $x$  is irrational and  $f(x) = \sqrt{2}/2$  if  $x$  is rational has no fixed point since  $0$  is rational and  $\sqrt{2}/2$  is irrational. Using the results about the global function in Appendix A this gives rise to a paradox. However, the paradox is not very interesting since it contains cycles. But we can easily remove the cycles and get a paradox with the same irrational flavor. To this end, let  $S = \{S_k\}_{k<\omega}$  be a set of sentence names. For each  $k < \omega$  define a function  $h_k : \mathbb{V}_S \rightarrow [0, 1]$  by letting  $h_k(v)$  be the binary decimal  $0.v(S_{k+1})v(S_{k+2})v(S_{k+3}) \dots$ . Now for each  $k < \omega$  define a function  $g_k : \mathbb{V}_S \rightarrow \{0, 1\}$  as follows. For  $v \in \mathbb{V}_S$ , let

$$g_k(v) = \begin{cases} 1 & \text{if } h_k(v) \in \mathbb{Q} \text{ and the } k\text{-th digit of the binary decimal form of } \frac{\sqrt{2}}{2} \text{ is } 1, \\ 0 & \text{otherwise.} \end{cases}$$

Now by Lemma 16, for each  $k < \omega$  we have  $\gamma_k \in S^+$  involving no element of  $\{S_0, S_1, \dots, S_k\}$  such that  $\gamma_k(v) = g_k(v)$  for each  $v \in \mathbb{V}_S$ . Let  $d$  be a denotation assignment on  $S$  such that  $d(S_k) = \gamma_k$ .

We claim that  $(S, d)$  is paradoxical. Assume (to reach a contradiction) that we have a truth-value assignment  $v \in \mathbb{V}_S$  which is acceptable on  $S$  with respect to  $d$ . Let  $y = 0.v(S_0)v(S_1)v(S_2) \dots$ . Note that if  $y \in \mathbb{Q}$ , then  $h_k(v) \in \mathbb{Q}$  for all  $k$  and if  $y \notin \mathbb{Q}$ , then  $h_k(v) \notin \mathbb{Q}$ . Now,  $v$  is acceptable, so  $v(S_k) = d(S_k)(v) = \gamma_k(v) = g_k(v)$ . Hence  $y = 0.g_0(v)g_1(v)g_2(v) \dots$ . If  $y \in \mathbb{Q}$ , then  $g_k(v) = 1$  if and

only if the  $k$ -th digit of the reduced binary form of  $\frac{\sqrt{2}}{2}$  is 1. Thus  $y = \frac{\sqrt{2}}{2} \notin \mathbb{Q}$ . This is a contradiction. Thus we must have  $y \notin \mathbb{Q}$ . But then  $g_k(v) = 0$  for all  $k$ , so  $y = 0 \in \mathbb{Q}$ . Again this is a contradiction.

## Appendix D: $\mathcal{F}$ -Systems

It is instructive to apply our terminology to the type of sentence systems that have been investigated most in the literature—we call these  $\mathcal{F}$ -systems.<sup>28</sup> Intuitively,  $\mathcal{F}$ -systems are sentence systems which are restricted in such a way that all the sentences can only *say* that other sentences in the system are *false*. For [2] the language  $L_P$  and his denotation function  $\delta$  give rise to an  $\mathcal{F}$ -system, since the only well-formed sentences of  $L_P$  are (possibly infinite) conjunctions of negations.<sup>29</sup> The motivation for theorists to restrict their attention to the reference structures inherent in  $\mathcal{F}$ -systems, we take it, is because both the Liar paradox and Yablo's paradox can be represented by  $\mathcal{F}$ -systems.<sup>30</sup> But there are many paradoxical systems such as Jourdain's and Curry's, which are not  $\mathcal{F}$ -systems—and many hypodoxical systems such as the Truth-teller, which are not  $\mathcal{F}$ -systems. For these reasons we have focused on the more inclusive language  $\mathcal{L}_S$  and the general class of sentence systems.  $\mathcal{F}$ -systems, however, are a subset of the general class of sentence systems discussed throughout this essay, defined as follows.

**Definition 31** Let  $G$  be a sink-free directed graph. The pair  $\mathcal{F}_G = (V(G), d)$  where  $d(x) = \bigwedge_{y \in N^+(x)} \neg y$  for each  $x \in V(G)$  is called the  $\mathcal{F}$ -system on  $G$ . Note that by construction  $\mathcal{G}_{V(G), d} = G$ .

It turns out that  $\mathcal{F}$ -system paradoxicality and hypodoxicality can be characterized in graph-theoretic terms.

**Definition 32** Let  $G$  be a directed graph. We call  $A \subseteq V(G)$  *independent* if  $G[A]$  is edgeless.

<sup>28</sup>We should note that we started this project in 2006 completely oblivious to the fact that there was any literature relating graph theory to the reference relations involved in paradox—except for the brief discussion in [14]. That paper in conjunction with the issue raised in [13] was the impetus for this project. We faintly recall being informed of [2] in 2008 by Andy McGonigal (during a late night conversation at the Phoenix). We have since adopted some of Cook's terminology (e.g. "denotation assignment" and "sentence name") and the overall presentation has benefited by comparing and contrasting his presentation with our own. We include this appendix to make note of the relations between that paper and this one.

<sup>29</sup>Note that  $L_P$  is not functionally complete in the way our  $\mathcal{L}_S$  is (see Section 1.3).  $L_P$  has conjunction, a class of sentence names  $S = \{\alpha_i\}_{i \in I}$ , a falsity predicate  $F$  and the only well-formed sentences in  $S^+$  are (unrestricted) conjunctions of the form  $\bigwedge_{i \in I} F(\alpha_i)$ . So, clearly, there is a function  $g$  from  $\mathbb{V}_S$  to  $\{0, 1\}$  such that there is no sentence  $\zeta_g \in S^+$  such that  $\llbracket \zeta_g \rrbracket = g$ .

<sup>30</sup>As pointed out by the anonymous referee, another nicety of the  $\mathcal{F}$ -system setup is that a graph is either paradoxical, determinate, or indeterminate (in Cook's terminology); whereas in our setup any dangerous graph is precarious as well.

**Definition 33** Let  $G$  be a directed graph. A *kernel* in  $G$  is an independent set of vertices  $K \subseteq V(G)$  such that each vertex in  $V(G) - K$  has an edge into  $K$ .

**Definition 34** Let  $X$  be a set. For any  $A \subseteq X$ , we call the function  $1_A : X \rightarrow \{0, 1\}$  given by

$$1_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

the *characteristic function* of  $A$  on  $X$ .

**Lemma 26** (Cook) *Let  $G$  be a sink-free directed graph. Then there is a bijection  $h$  between the kernels of  $G$  and the acceptable truth assignments on  $\mathcal{F}_G$  given by  $h(K) = 1_K$ .*

*Proof* We first need to show that  $h$  maps kernels to acceptable truth assignments. So, let  $K$  be a kernel and let  $v = h(K)$ . Then for any  $x \in V(G)$  we have

$$\llbracket d(x) \rrbracket(v) = \llbracket \bigwedge_{y \in N^+(x)} \neg y \rrbracket(v) = \bigwedge_{y \in N^+(x)} \neg v(y).$$

Since  $K$  is independent, if  $x \in K$ , then  $N^+(x) \subseteq V(G) - K$  and hence we have  $\llbracket d(x) \rrbracket(v) = 1 = v(x)$ . Since each vertex in  $V(G) - K$  has an edge into  $K$ , if  $x \in V(G) - K$  we have  $\llbracket d(x) \rrbracket(v) = 0 = v(x)$ . Thus  $v$  is acceptable on  $V(G)$ .

Next we check that  $h$  is injective. So, let  $K_1, K_2$  be kernels in  $G$  such that  $h(K_1) = h(K_2)$ . Then  $1_{K_1} = 1_{K_2}$  and hence  $K_1 = K_2$ . Thus  $h$  is injective.

It remains to check that  $h$  is surjective. So, let  $v$  be an acceptable truth assignment on  $\mathcal{F}_G$ . Put  $K = \{x \in V(G) \mid v(x) = 1\}$ . Since  $v$  is acceptable, we see that  $K$  must be independent. Now, pick  $x \in V(G) - K$ . Since  $v(x) = 0$  and  $G$  is sink-free, for some  $y \in N^+(x)$  we must have  $v(y) = 1$  and hence  $y \in K$ . Thus  $K$  is a kernel in  $G$ . By definition we have  $h(K) = v$ . Hence  $h$  is surjective. □

Lemma 26 immediately implies Cook’s graph theoretical characterization of  $\mathcal{F}$ -system paradox.

**Theorem 27** (Cook) *Let  $G$  be a sink-free directed graph. Then  $\mathcal{F}_G$  is paradoxical if and only if  $G$  has no kernel.*

Additionally, we get a graph theoretical characterization of  $\mathcal{F}$ -system hypodox.

**Theorem 28** (Cook) *Let  $G$  be a sink-free directed graph. Then  $\mathcal{F}_G$  is hypodoxical if and only if  $G$  has more than one kernel.*

Since a directed graph  $G$  is dangerous (precarious) if  $\mathcal{F}_G$  is a paradox (hypodox) we get the following corollaries.

**Corollary 29** *If  $G$  is a sink-free directed graph with no kernel, then  $G$  is dangerous.*

**Corollary 30** *If  $G$  is a sink-free directed graph with more than one kernel, then  $G$  is precarious.*

Yablo [14] gave some sufficient conditions for an  $\mathcal{F}$ -system to be paradoxical. In light of Cook's theorem above we can view these as sufficient conditions for a sink-free directed graph to have no kernel. Here we give a generalization of Yablo's conditions.

**Definition 35** Let  $A$  be an infinite set. We say that  $B \subseteq A$  is *cofinite* in  $A$  if  $A - B$  is a finite set.

**Lemma 31** *Fix  $n \geq 1$ . Let  $G$  be an acyclic sink-free directed graph such that for any  $n$  different vertices  $x_1, x_2, \dots, x_n \in V(G)$  the set  $\bigcup_{1 \leq i \leq n} N^+(x_i)$  is cofinite in  $V(G)$ . Then  $G$  contains no kernel.<sup>31</sup>*

*Proof* Assume the lemma is false and let  $K$  be a kernel in  $G$ . Since  $G$  is acyclic and sink-free, it must be infinite. Also, since  $G$  is acyclic, Lemma 20 gives us a topological sort  $<$  on  $V(G)$ .

First assume  $K$  is finite. Then we have  $z \in V(G)$  such that for each  $x \geq z$ ,  $x \notin K$ . But then since  $z \notin K$ , we must have  $y \in K$  such that  $zy \in E(G)$  and hence  $y > z$ . This is a contradiction.

Hence  $K$  is infinite. Thus we may choose different  $x_1, x_2, \dots, x_n \in K$ . By hypothesis,  $D = \bigcup_{1 \leq i \leq n} N^+(x_i)$  is cofinite in  $V(G)$ . But since  $K$  is independent,  $K \subseteq V(G) - D$  and hence  $K$  is finite. This final contradiction completes the proof.  $\square$

The case  $n = 1$  gives Yablo's condition.

**Corollary 32** (Yablo) *Let  $G$  be an acyclic sink-free directed graph such that for every  $x \in A$  the set  $N^+(x)$  is cofinite in  $V(G)$ , then  $G$  contains no kernel.*

Lemma 31 is more powerful than Yablo's condition which can be seen by considering the following example. Let  $p_i$  denote the  $i$ -th prime number, so  $p_0 = 2$ ,  $p_1 = 3$ ,  $p_2 = 5$ , etc. Let  $G$  be the directed graph with vertex set  $\mathbb{N}$  and an edge from  $a$  to  $b$  if and only if  $a < b$  and  $b \neq p_a^n$  for any  $n \in \mathbb{N}$ . That is,  $a$  has an edge to every natural that is not a power of the  $a$ -th prime number. Since there are infinitely many powers of each prime number, no vertex in  $G$  has cofinite out degree. Hence Yablo's condition does not apply. But for any  $a, b \in \mathbb{N}$  with  $a \neq b$  and  $m, n \geq 1$  we have  $p_a^m \neq p_b^n$  and hence  $N^+(a) \cup N^+(b)$

<sup>31</sup>It is actually enough to assume there is some  $A \subseteq V(G)$  which is cofinite in  $V(G)$  such that for any  $n$  different vertices  $x_1, x_2, \dots, x_n \in A$  the set  $\bigcup_{1 \leq i \leq n} N^+(x_i)$  is cofinite in  $V(G)$ .

is cofinite in  $V(G)$ . Thus we may apply Lemma 31 to conclude that  $G$  has no kernel and hence  $\mathcal{F}_G$  is paradoxical.

We can also use our general necessary conditions for a graph to be dangerous to conclude that certain directed graphs must contain kernels.

**Corollary 33** *Let  $G$  be a sink-free directed graph. Each of the following is a sufficient condition for  $G$  to contain a kernel.*

- *The underlying undirected graph of  $G$  is acyclic.*
- *$G$  is acyclic and only finitely many vertices have infinite out degree.*

## References

1. Bruijn, N.G., & Erdos, P. (1951). A colour problem for infinite graphs and a problem in the theory of relations. *Indagationes Mathematicae*, 13, 371–373.
2. Cook, R.T. (2004). Patterns of paradox. *Journal of Symbolic Logic*, 69(3), 767–774.
3. Cook, R.T. (2006). There are non-circular paradoxes (But Yablo's isn't one of them!). *The Monist*, 89(1), 118–149.
4. Diestel, R. (2010). *Graph theory* (4th ed.). Heidelberg: Springer.
5. Eldridge-Smith, P. (2008). *The liar paradox and its relatives*. PhD thesis, The Australian National University.
6. Herzberger, H. (1970). Paradoxes of grounding in semantics. *The Journal of Philosophy*, 67(6), 145–167.
7. Kripke, S. (1975). Outline of a theory of truth. *Journal of Philosophy*, 72(19), 690–716.
8. Priest, G. (1997). Yablo's paradox. *Analysis*, 57(4), 236–242.
9. Rabern, L., Rabern, B., Macauley, M. (2012). Dangerous reference graphs and semantic paradoxes. *Journal of Philosophical Logic*. doi:10.1007/s10992-012-9246-2.
10. Schlenker, P. (2007). The elimination of self-reference: generalized Yablo-series and the theory of truth. *Journal of philosophical logic*, 36(3), 251–307.
11. Yablo, S. (1982). Grounding, dependence, and paradox. *Journal of Philosophical Logic*, 11, 117–137.
12. Yablo, S. (1985). Truth and reflection. *Journal of Philosophical Logic*, 14, 297–349.
13. Yablo, S. (1993). Paradox without self-reference. *Analysis*, 53(4), 251–252.
14. Yablo, S. (2006). Circularity and paradox. In S.A. Pedersen, T. Bolander, V.F. Hendricks (Eds.), *Self-reference* (Chap. 8, pp. 139–157). CSLI Press.