Doctor's Diagnosis Sustained[1]

Albert A. Johnstone

I.

I thank Professor Roy Cook for his energetic response[2] to my article, "The Liar Syndrome". It is flattering to find oneself capable of provoking such a high degree of passion, even if unremittingly negative in tenor, just as it is reassuring to find the corpus of one's account standing firm in the face of such a vigorous assault.

Before responding in detail to the barrage of criticisms, I should like first to underscore two points about my article. If, for logicians, the article does not quite proceed in the manner to which they are accustomed, this is not simply because it is intended for a somewhat wider audience. Rather it is because the source of the Liar and related paradoxes lies in the meaninglessness of certain types of statement. The logician's preferred method of testing hypotheses, that of deriving contradictions (a preference illustrated by Cook's advocacy of the Tarski Biconditional, p. 7), simply will not work when it is a matter of ferreting out meaningless statements. More often than not, nonsense generates no contradiction. What is required is a good nose. In formal systems, the task of excluding meaninglessness devolves on the formation rules of the system as dictated by the relevant olfactory findings. Certain combinations of symbols are excluded because they make no sense on their intended interpretation. One of the central theses of my article is that certain meaningless statements, because they look like genuine statements, have been unwittingly admitted into formal systems. Such pseudo-statements are also found in English where, fortunately, speakers generally dismiss them as irrelevant or nonsensical despite their grammatical correctness. They could and should be likewise dismissed from formal systems. Instead, because they conform to existing formation rules, they are widely assumed to make sense. The unfortunate consequence of this erroneous assumption has been an interminable search for some tolerable, catastrophe-preventing conceptual artifice,[3] and also, of course, the espousal of Gödel's Theorem.

[1] My appreciation is extended to Editor Steen Brock whose activities have obliged me to add needed clarification to a number of points made in "The Liar Syndrome."

[2] In his article, "Curing the Liar Syndrome," in the present issue of *Sats*.

[3] Gupta and Belnap's revision theory of truth that Cook recommends (p. 7, note 5) is one of the more recent constructions of the sort, impressive but arbitrary.

The second point to underscore concerns the sense I give to the term 'definition'. A central claim of my article is that the cases of meaninglessness being considered are caused by definitional circularity (*not* circularity of reference, as Cook mistakenly has me say (p. 1)), which circularity generates the three symptoms of the Liar syndrome (*not* the other way around following Cook, who mistakenly makes circularity the symptom, and the Liar syndrome the cause (p. 1)). It should be clear from the article that the term 'definition' implicit in the expression 'definitional circularity' is being given two distinct senses, neither of which is that of a dictionary definition. These two senses figure respectively in the discussions of the two types of cases the article considers: *self-referential statements* and *criterially circular predications*. In the discussion of self-referential statements, I propose speaking of statement-designators as having a definition, where the definition of the designator states what statement the speaker intends to designate with the designator. In the discussion of predication, the relevant definitions are clearly individualized ones, definitions that state the necessary and sufficient conditions for the possession of a given property by a particular individual. The reason for the individualization is that the cases with which the article is concerned are ones in which the usual criteria for possession of the property turn out to be circular when applied to a particular individual, with the result that the predicate is undefined for that individual. The resulting situation is similar to one in mathematics where a function is defined for certain values of the variable, but undefined for others.

II.

In my article I argue that circular definition (in the two senses mentioned) generates the Liar syndrome, and, in certain circumstances, paradoxes (the Liar and kin). At the outset (p. 1), Cook makes two general objections to my conclusions. He claims on the one hand that there exist cases of circularity where there is no Liar syndrome, and on the other that there exist cases of paradox where there is no circularity and no Liar syndrome. According to Cook, my analysis consequently turns out to be both faulty and inadequate. The first of his two claims, if substantiated, would cast considerable doubt on my analysis. Some plausible reason might, of course, be found to explain the existence of the healthy cases of circularity, a possibility Cook seems not to have considered. The second of Cook's two claims is in fact irrelevant. I nowhere claim to present a God's eye view of the source of all known paradoxes, let alone all possible ones. Indeed, it is far from obvious that there is only one such source of paradox. The account I present concerns a certain group of paradoxes (Liar and kin), and should stand or fall on its specific merits. If it is plausible and convincing, it is not invalidated by its failure to explain paradoxes afflicting the other side of town.

In the course of his criticism Cook proffers one example in support of each of his two claims (pp. 6, 11). He adds a third less dramatic claim to the effect that there exist cases of self-reference in which there is no circular definition, in support of which claim he likewise adduces an example (p. 5). We shall examine in turn each of these three putative counterexamples to my account, beginning with the third. Curiously enough, all three of Cook's examples turn out to be misnamed; the proposed Liar is not a Liar, the Truth is not a truth, and the Paradox is not a paradox.

To illustrate his claim that there can be self-reference without circular definition, Cook introduces Quine's version of the Liar. The proposed counterexample runs as follows: "appended to its quotation yields a falsehood" appended to its quotation yields a falsehood. Cook argues that the self-reference is generated by a definite description, and that no definition – "stipulational assistance" (p. 6) – is involved. He is, of course, quite right on the latter point. There is no definition of the sort involved. However, Cook is doubly mistaken on the first point. To begin with, Quine's paradox is not a version of the Liar at all. It deals in predicates applied to their own quotations, and consequently should be classed with the second group of statements considered in my article, criterially circular predications. In point of fact, Quine's paradoxical statement is simply the Grelling statement decked out with quotation marks and talk of quotations. In cases of the sort, the relevant definition concerns the criteria (necessary and sufficient conditions) for the application of a predicate to a particular individual. As argued in my article, in the Grelling statement the criteria for the application of the predicate to itself are circular. Unfortunately, Cook chooses to ignore this particular section of my article. Consequently, he misinterprets the sense in which 'definition' is to be understood in such cases. Naturally enough, he fails to find any definitional circularity of the relevant sort because he is looking for something else, namely, "stipulational assistance". In his subsequent discussion of Gödel, Cook repeats the mistake, and once again misses the point with his vigorous protests that no stipulational definition is involved. My main complaint with the Gödel sentence is that, like the Grelling statement, it predicates of an individual a predicate that is undefined for that particular individual, and in this sense involves definitional circularity.

Let us proceed to Cook's claim that there exist cases of circularity that do not generate the Liar syndrome. The counterexample proposed by Cook (p. 6), which he terms "the Self-referential Truth", is a circular and self-referential statement that says of itself that either it is true or it is false. Cook finds this statement to be "intuitively ... just plain true", and consequently to provide a self-referential statement unafflicted with the Liar syndrome. Now, with all due respect to Cook's intuitions, my own intuitions find the Self-referential Truth to be nonsense, and not a truth at all. It gives a first impression of being true, but then so does the self-referential statement, this very statement is true. The impression is an illusion. In both cases it is impossible to say what the statement states without bringing in the statement itself as part of the proposed explanation of its meaning. The statement in both instances is incomplete, a pseudo-statement in which the subject term has no referent. In this regard, Cook's Self-referential Truth (so-called) resembles the pseudo-statement, the present king of France is either bald or not (or the pseudo-statement, either the present king of France is bald, or the present king of France is not bald). Both appear to enunciate something logically true, but in both cases the subject term fails to refer to anything, and the statement made fails to be a genuine one.

While Cook's third putative counterexample is irrelevant to the aim of my article, it is interesting to note that it also fails in the mission Cook assigns it. Yablo's Paradox is touted by Cook to be an instance of "patients who clearly suffer from the same disorder as the Liar, but who fail to display the symptom of circularity" (p. 11). The paradox features an infinite series of statements, each stating that all further statements in the series are false. Cook rightly points out that there is no circularity of reference, and "no assignment of truth and falsity such that all the relevant Tarski Biconditionals come out true". However, he fails to note that since each statement is about other members in an infinite series, each is vacuous. Consequently, it is only to be expected that there is no proper assignment of truth or falsity; it would be astounding if there were. Yablo's Paradox is in fact not a paradox at all. It may well create an awkward moment for people who play with infinite series, but it is not something with which any general account of paradox need be concerned.

With the failure of all three alleged counterexamples, Cook's three general objections to the account in my article are left without visible means of support. Cook suggests that many further counterexamples exist in the literature, but I find the suggestion unlikely. Consider, for instance, Anil Gupta's claim that self-reference figures without mishap in logical laws such as, "No sentence is both true and false".[4] Contrary to what Gupta suggests, there is no self-reference in such a statement. Logical laws make no explicit reference to themselves,

[4] Anil Gupta, 'Truth and Paradox', in *Recent Essays on Truth and the Liar Paradox*, ed. Robert L. Martin (Oxford: Clarendon press, 1984), pp. 175-23, p. 210.

and no such reference is even intended. Unless a statement attempts to refer to itself, no vacuity is generated. Gupta further argues that self-reference is allowed in certain types of everyday reasoning, such as in an example he gives (sometimes termed "Gupta's Puzzle" in the literature[5]). His somewhat complex example reduces essentially to the following situation: speaker A states that everything speaker B says is true; A also states that something B says is false; B states that at most one statement made by A is true. Gupta finds it unproblematic to reason that since A's two statements contradict each other, B's statement is true, and that consequently A's second statement is false and A's first statement true. However, reasoning of the sort is, on the contrary, highly problematic. The given statements have no content apart from their evaluations of each other. As a result they end up being vacuous evaluations for which it is impossible to say what is being said. The two statements made by A appear to contradict each other, but surely two utterances can contradict only if they say something, and the above statements, of course, merely appear to say something.

III.

We turn now to Cook's criticisms of the Liar syndrome proper with its three semantic disorders. It may be helpful here to consider briefly a simple model of the Liar syndrome. Let us suppose I consult the dictionary to learn the meaning of the noun, 'tat', and I am informed that it means a tall tat. The explanation given is not only uninformative, but it suffers from the three disorders of the Liar syndrome. It is semantically vacuous since the word to be explained figures in the proposed explanation. It is semantically absurd since the meaning of a noun is made semantically equivalent to the meaning of that same noun together with an adjective. The adjective in question is made cataleptic in the sense that it can no longer function as adjectives normally do. To say that something is tall normally implies that it could in principle not be tall. In the present case, a tat that was not tall would no longer be a tat. It is not just that there are in fact no short tats, but that in principle there can be no such thing. The definition of 'a tat' as 'a tall tat' has deprived the adjective 'tall' of its normal mode of functioning.

The important question for our purposes is whether the three disorders of the Liar syndrome are generated in analogous fashion by definitional circularity in self-referential statements and in predicative criteria.

As a generalizable instance of statemental self-reference, I introduce the

[5] See Gupta, 'Truth and Paradox', p. 176; Jon Barwise and John Etchemendy, *The Liar: An Essay on Truth and Circularity* (New York: Oxford University Press, 1987), p. 23.

statement, this very statement is true. A statement of the sort often elicits puzzlement followed by the question, 'What statement?' Alternative appropriate answers to the question would be 'this very statement', or 'this very statement is true', or the explanation that the statement meant by 'this very statement' is the statement, this very statement is true. For the purposes of discussion I formalize the explanation as follows, letting 'p' represent the statement-designator, 'this very statement', and letting '$=_{ds}$' mean 'is by designation' (as distinct from 'is by definition'):

(1) $p =_{ds} p$ is true

The formalization is a straightforward enough rendition of English idiom since the statement which 'this very statement' is being used to designate, is the statement, this very statement is true. Cook objects to the notation on the grounds that it is unnecessary and that it confuses use with mention (p. 4).

Both objections are in fact irrelevant on two counts. First, it is important to note that Cook himself apparently does not think that the confusion makes the self-referential attribution of truth any less semantically vacuous, since in due course (p. 6) he grants the existence of this first disorder. Ironically enough, the existence of vacuity is the central point at this stage of the article; discussion of the two further disorders of absurdity and catalepsy could be dropped without detriment to the main conclusions reached. Indeed, in limiting cases of circularity such as when 'p' is decreed to name the statement, p, vacuity alone is generated without absurdity or catalepsy. Let us continue nevertheless.

Second, the notation used in (1) is quite irrelevant to the substance of the point being made. Its intended purpose is clarity of exposition, not canonical reform. It is sufficient that it capture English usage according to which the designator, 'p', and the sentence 'p is true' that expresses the designator's referent, are intersubstitutable (with certain limitations). The expressions in (1) on the two sides of '$=_{ds}$' may be used interchangeably in answer to the question, 'What statement?' In this regard, (1) is a formalization of a quite usual English practice. An instance of such practice might be a case where I declare, "That statement is false", in response to the claim, "John stole it", and a third party wonders what statement I mean, and so asks, "what statement is false?" In answer I could simply say "that statement", or more helpfully say "John stole it", or I might explain how I was using the statement-designator.

The crucial point is not the notation but the fact that the two intersubstitutable expressions in (1) both contain the same statement-designator, 'p'. That designator has exactly the same referent in both instances of its occurrence (and in both cases 'p' is being mentioned rather than used). Since the designators in the two expressions are the same, and the two expressions are intersubstitutable for semantic reasons (as distinct from logical ones), the two expressions must be semantically identical or equivalent, that is, identical or equivalent in meaning. An absurdity thereby results since the second expression contains a predicate over and above the designator. Consequently, the definition (explanation) is in fact ruling the first expression to be semantically identical with itself combined with a predicate. Since the predicate is meaningful, the semantic situation created is absurd. The point is more obvious when the predicate is one such as 'false' or 'possible', but it holds nevertheless of 'true'.

Notation clearly plays no essential role in establishing the above conclusions – contrary to Cook's claim that it does (p. 8). However, since Cook attaches so much importance to the notation used, we should perhaps look more closely at the issue. Cook objects to my notation on the ground that "identity is a two place relation taking terms as argument" (p. 5). Presumably Cook thinks that a statement is not rightly treated as a term. However, he gives no reason whatever in support of his pronouncement. Since statements purport to state facts, and since it is far from obvious that there are no such things as facts, I see no reason not to treat statements as terms, and so use my notation. In place of the latter, Cook proposes a notation (p. 5) in which the subscript, 'ds', is replaced by 'df', and quotes are placed around the right-hand side of the identity. Cook's proposal is an odd one to make since such a correction is clearly incorrect on a straightforward reading of it. It makes the statement-designator designate a sentence instead of a statement, and thus perverts speaker intent. In support of his formulation Cook claims that the expressions on the two sides of the identity symbol "do not differ grammatically" (p. 5), but his claim is patently false. One expression means a particular statement, the other a string of words. Cook also appeals to what "we" do in such cases, but "we" refers in fact to no more than a certain school of logicians. Likewise, he touts his notation as "time-honored," but the characterization is hardly relevant to logical correctness – especially in the present context where the aim of my article is precisely to question certain other 'time-honored' assumptions in formal logic. Cook nowhere attempts to explain what he means by a use/mention confusion, or why he thinks that in order to avoid the confusion, quotes are mandatory around statements in certain contexts.

When Cook comes to dispute the further disorder of semantic absurdity, he claims that whereas I am misled by my notation and the use/mention confusion, with his reformulation of (1), the mistake is harder to make (p. 8). According to Cook, the latter "is merely an assertion that two distinct names of statements in fact have the same reference" (p. 8). Now, surely this claim is itself a confused one to make since the expression on the left hand side has no quotes, and so is not a name, while the expression on the right hand side is a sentence, and a sentence is not a name of anything. More importantly, Cook's confusion of a sentence with a name leads him to assimilate mistakenly the situation under discussion to one in which two proper names or identifying descriptions have the same referent, as is the case, for instance, with 'Cicero' and 'Tully', or with 'Ortcutt' and 'the man in the brown hat'. In cases of the sort the expressions involved may be intersubstitutable (in transparent contexts) and nevertheless differ semantically. In the case under consideration, the situation is quite different. As noted above, the first of the two intersubstitutable expressions contains the same designator as the second, while the second contains in addition a semantic predicate. Ironically, Cook is being misled by the very notation he recommends as less likely to mislead.

From the above conclusions it also follows that the predicate, true, as used in (1) is cataleptic. The predicate, true, purports to modify the statement, p. Yet, since p is semantically identical with the statement, p is true, the predicate is also an integral part of the statement, p. For this reason it cannot function as a normal predicate. If the predicate were other than the predicate, true, (the possibility of which is a proper feature of predicates), the statement would no longer be the statement, p. In the situation created by the definition, the predicate cannot function as it must if it is to be a proper predicate.

To sum up, since (1) is representative of self-referential statements with semantic predicates, it must be concluded that the Liar syndrome is alive and comfortably installed in semantically self-referential statements.

IV.

With the discussion of Gödel's Theorem, Cook's criticisms of my article reach a veritable paroxysm of misunderstanding and misrepresentation. As noted earlier, Cook rather strangely ignores completely the discussion of criterially circular predications. If the latter were indeed irrelevant to understanding the Gödel sentence on which the theorem rests, and if an awareness of semantic self-referential statements were sufficient, it would be perverse on my part to inflict an additional tortuous topic on the reader before proceeding to a discussion of Gödel's Theorem. The discussion of criterial circular predication is necessary, unfortunately, since the Gödel sentence embodies *both* criterially circular predication *and* circular statemental self-reference.

To illustrate the situation, consider definition (7) of my article.

(7) $Nn \equiv_{df} \sim P(Nn)$

The definition is an individualized one, and states the necessary and sufficient conditions for a particular individual, n, to have a particular property, N. Those conditions are that the statement, n has the property N, not be provable. As argued in my article, the criteria for possession of the property are circular since they define n's having of N in terms of n's having of N. As a result the criteria are vacuous; they postulate a semantic absurdity (that Nn should be semantically equivalent to itself conjoined with a semantic predicate); they render cataleptic the predicate, 'not provable'. In sum, they are infected with the Liar syndrome.

In virtue of its peculiar form, definition (7) admits of an additional reading. As with any definition so formulated, the right-hand side explains what the left-hand side is saying. Hence (7) may be interpreted as saying that the statement, Nn, says of itself that it is not provable. On such an interpretation of its definition, the statement is semantically self-referential, and, of course, as such is afflicted with the three disorders of the Liar syndrome.

These two features of the statement defined in (7) are shared by the statement made by the Gödel sentence on its intended interpretation. The fact comes out most clearly in the metalinguistic description that Gödel gives of the Gödel sentence in his informal presentation of the theorem that precedes the fuller version.[6] Here Gödel defines a class, K, of class numbers (Gödel numbers assigned to classes) as the class of numbers such that the sentence stating that they belong to the class they number is not provable. A number is assigned to the class, K, and the Gödel sentence states that the number belongs to the class it numbers. In Gödel's metalinguistic notation, the definition of the Gödel sentence may be represented roughly as follows, where 'q' represents the class-number of the class, 'R(q)' the class-sign of the class, '[R(q);q]' the formula that results from substituting the class-number, 'q', into the class sign, 'R(n)' with class number n, and where '~Bew' is to be read as 'not provable':

(15) $[R(q);q] \equiv \sim Bew[R(q);q]$

Clearly (15) mirrors (7) – except for the fact that it deals in sentences rather than statements made by sentences. My article assumes somewhat naively that the resemblance will be obvious to aficionados of Gödel's Theorem. The similarity in structure implies that arguments that apply to (7) carry over to

[6] See Kurt Gödel, 'On formally undecidable propositions of *Principia mathematica* and related systems I', in *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, ed. Jean van Heijenoort (Cambridge, Mass.: Harvard University Press, 1967), pp. 596-616, p. 598.

(15), and that the Gödel sentence as defined in (15), like the statement defined in (7), is both a criterially circular predication, and, on its intended interpretation, a semantically self-referential statement. Consequently, there are two good reasons (the two reasons given on pages 51 and 52 of my article) for affirming that the Gödel sentence is syndrome-infected and cannot make a genuine statement.

Cook's objections sail past most of this. Cook misinterprets my article as claiming that Gödel introduces "a new predicate", that Gödel adds "vocabulary to first-order arithmetic", and that Gödel defines the Gödel sentence "to be identical to the statement that G is not provable" (pp. 13, 14). I quite agree that Gödel does none of this. What he does do is allow a substitution that generates a criterially circular and hence meaningless predication. He mistakenly assumes that the Gödel sentence, because it results from the substitution of a particular natural number sign for a free variable in a particular class sign (characterized in my article as 'a derivability predicate'), is meaningful on its intended interpretation. The assumption is of a piece with his conviction concerning the formal system he uses, of which he states that "it is easy to state with complete precision *which* sequences of primitive signs are meaningful formulas and which are not".[7] Cook's response to my questioning the correctness of Gödel's assumption is to declare: "Fortunately for Gödel, the details of his proof guarantee that such an assumption is unproblematic" (p. 13). However, the details of the proof are not the issue: Gödel's assumption concerns the substitution rules, and when they are examined, they show precisely the contrary.

The result of the substitution allowed by Gödel is a sentence that is not only a criterially circular predication, but also one that may be interpreted as saying of itself that it is not provable. Gödel himself interprets it in these terms when he declares: "We therefore have before us a proposition that says about itself that it is not provable [in PM]".[8] In light of Gödel's own remarks, Cook's claim that "Johnstone has failed to demonstrate that the Gödel sentence refers to itself" (p. 12), makes no sense. Furthermore, the question of what allows Gödel to characterize the Gödel sentence as referring to itself is one that should have an answer. On Cook's account of the Gödel sentence as "pure arithmetic" or "just an arithmetical claim about the natural numbers" (p. 15), Gödel's characterization is aberrant and inexplicable. On the above account, it is a natural one for Gödel to give.

In point of fact, it is quite misleading to describe the Gödel sentence as

[7] Ibid., p. 597.

[8] Ibid., p. 598, last paragraph.

"pure arithmetic". Admittedly, it is a sentence in the object language, which, in the present case is roughly the formal system of *Principia Mathematica* supplemented with Peano's axioms and constants of arithmetic. However, the sentence incorporates a metalinguistic predicate. Gödel defines the notions of 'a formula resulting from an axiom schema', 'an immediate consequence', and 'a provable formula', in terms of the machinery available in the formal system being used.[9] These notions are all used to make statements about statements, and, whereas statements about numbers are arithmetic, statements about statements about numbers are metalinguistic – a distinction immediately apparent to anyone familiar with the use/mention distinction. Since the Gödel sentence concerns numbers whose membership in a particular class is not provable, the sentence is not purely arithmetic.

In the final analysis, the Gödel sentence is not decidable because it is formed through a substitution that produces circular definition of predicate criteria. In this regard it is similar to the statement defined in (7) in terms of class-numbers of which it is not provable they are members of the class they number. There are no criteria (non-arbitrary ones) for determining the truth or falsity of the statement being made. Both sentences suffer from the Liar syndrome (actually a double dose), and fail to make genuine statements on their intended interpretation.

In section IV of my article, it was found that the sentences used to make self-referential statements lead a double life; they are ambiguous. The same may be said of sentences formed with circularly defined semantic predicates. Since the Gödel sentence is both self-referential and predicatively circular, it also leads a double life. The same sequence of symbols may serve to make either a nonsensical statement or a genuine statement (true or false) about the nonsensical statement. While the Gödel sentence (represented metalinguistically as '~Bew[R(q);q]') makes a nonsensical statement about itself, the same sequence of symbols could be used to say that the Gödel sentence is not derivable, hence used to make a statement that is true. The ambiguity of the notation is what allows Gödel to remark of the Gödel sentence that "[R(q);q] is true, for [R(q);q] is indeed unprovable (being undecidable)".[10] He is able to claim, plausibly enough, that the sentence states something true since he does not realize that two distinct statements are involved.

Once again, Cook's criticisms sail past these complexities. Indeed, my report of Gödel's claim (that the Gödel sentence states something true) is summarily

[9] Ibid, p. 606, definitions 38 to 46.

[10] Ibid, p. 599. See also a second passage to this effect, p. 615, footnote 67.

dismissed by Cook as "historically absurd" (p. 14, note 7) – despite it's being supported by two passages in Gödel's text, the above quote and footnote 67, page 615. Historical revisionism of the sort can only result in casting substantive doubt on its promoter's scholarship.

It should be clear that any formal system containing semantic predicates is bound to encounter problems if it admits sentences with statemental self-reference and criterially circular predication. Any such system needs not only a third truth-value in order to accommodate the pseudo-statements made, but also some means of identifying and disambiguating sentences of the sort. If it has the former but not the latter (as in the Intuitionism featured in Cook's example (p. 14, note 7)), then the Gödel sentence will be treated as meaningful, and Gödel's Theorem will be incorrectly considered to be valid. The source of the problems must be not only sought, but duly recognized when found.

To sum up, the doctor's diagnosis is sustained, and the second opinion denied.

Department of Philosophy
University of Oregon
bertjohnst@yahoo.com