An open letter to Frank Bäuerle

Introduction

In your logic class last Friday —or maybe it was two Fridays ago— , you discussed some famous paradoxes, in particular:

- the barber paradox,
- Russell's paradox, and
- the 'un-self-descriptive' paradox,

all of which are somewhat similar. You claimed that the nature of the paradoxes had something to do with self-reference. I did not understand this comment, but felt that it would be imprudent to delve further into the matter, seeing as how we were off-topic in the first place. Now that I have had the opportunity to focus and refine my thoughts, I would like to share them with you. Whence this letter.

My position

In brief, my position is that a 'paradox' arises precisely when one assumes **false** . (Or equivalently, something that implies **false** .)

A parenthetical remark. Of course, proving my position would amount to proving the consistency of human logic, but that is not my aim. Humans can, do, will, and must use human logic, whether it is consistent or not, and so I feel that nothing is to be gained by questioning its consistency. My aim here is only to provide a way of understanding paradoxes *within* the realm of human logic and understanding.

(**End** of A parenthetical remark.)

I am aware that the meaning I have given to the term 'paradox' does not coincide with the usual one. Indeed, in common usage, we say that we have encountered a paradox when we feel that *valid* reasoning has led us to a contradiction: for example, when "logic" tells us to go left and also right, or that some value is simultaneously 0 and 1 , or that the barber shaves himself and also doesn't. A situation is paradoxical when a contradiction seems to have entered the world of the living!

My claim, however, is that these situations are entirely understandable: we have simply assumed **false** without realizing it. Certainly we *feel* that the situation is paradoxical, but that is only because we are not aware of our contradictory assumptions. And it is all too easy to get into such a situation, because natural language and natural reasoning are extremely fuzzy.

It follows from my opinion that paradoxes have nothing to do with self-reference *per se* . At the same time, I admit that self-referential language appears to cater to the easy formulation of seeming paradoxes. I return to this point at length in the final section.

Also, we will see that several paradoxes seem paradoxical only because we have not followed George Boole in stripping **true** and **false** of their philosophical connotations, and treating them simply as values in a mathematical domain.

In the following three sections, I investigate the paradoxes mentioned in the introduction.

The barber paradox

The barber paradox involves the following scenario: "In a certain village, the barber shaves precisely the villagers who do not shave themselves." . Does the barber shave himself? If he does, then by definition he doesn't, and if he doesn't, then by definition he does. This appears to be a paradox.

As always in human investigation, there can be rational debate as to the cause of the contradiction.

We might decide that human logic itself is flawed. But that conclusion is unacceptable to me, because I have never met a single person who lives without human logic. (More to the point: I refuse to live without it.)

Rather, I claim that the source of the contradiction is a hidden linguistic assumption.

When we use definite noun phrases like 'the barber' as above, it is tacitly assumed that there is a barber in the village. (For example, unless I am lying or joking, it is inappropriate to say "My son likes reggae." if I don't have a son. Even though I don't say it explicitly, my statement implies that I have (precisely!) one son.)

Put mathematically, then, we have assumed:

$$(0) \qquad \langle \exists x :: \ \text{barber}.x \ \rangle \qquad ,$$

where x is of type 'villager' . The scenario provides another assumption:

$$(1) \qquad \text{barber}.x \quad \Rightarrow \quad \langle \forall y :: \ (x \text{ shaves } y) \ \equiv \ \neg(y \text{ shaves } y) \ \rangle \qquad ,$$

where y is of type 'villager' . (Hopefully my notational conventions are conspicuous enough!) Since we have agreed to use the principles of human logic, we may reason as follows:

$$\begin{array}{ll} & \text{barber}.x \\ \Rightarrow & \{ \ (1) \ \} \\ & \langle \forall y :: \ (x \text{ shaves } y) \ \equiv \ \neg(y \text{ shaves } y) \ \rangle \\ \Rightarrow & \{ \text{ instantiation, since } x \text{ is of type villager } \} \\ & (x \text{ shaves } x) \ \equiv \ \neg(x \text{ shaves } x) \\ \equiv & \{ \text{ predicate calculus } \} \\ & \textbf{false} \qquad , \end{array}$$

whence we conclude:

(2) $\neg\,\text{barber}\,.x$.

The conjunction of (0) and (2) implies **false** . Hence our assumptions imply **false** . There is no paradox, only contradictory assumptions.

As we saw, the feeling of paradox stemmed from a tacit linguistic assumption (0) . Without this assumption, the scenario contains only assumption (1) , from which we may conclude that there is no barber in the village, a perfectly sensible and logical conclusion. We arrive at a contradiction only when we assume the negation of this conclusion, namely, that there *is* a barber in the village.

Russell's paradox

The mathematicians Gottlob Frege and Georg Cantor hoped that a set could be defined naively as a collection of distinct thought objects. Bertrand Russell (apparently) dashed their hopes by considering the following scenario: "Let S be the set of all sets which do not contain themselves." . Does S contain itself? If it does, then by definition it doesn't, and if it doesn't, then by definition it does. Egads! Another paradox!

Whoa, Nellie. Before we go nuts, let us observe that Russell's scenario is teeming with hidden assumptions. First of all, he tacitly assumes the consistency of human logic —as anyone who makes a logical argument does— , but I am not about to quibble with him there. Second of all, as in the barber paradox, he tacitly assumes that S exists. Third of all, he assumes that S is a set. And finally, he assumes that the 'contained in' relation is defined for all sets.

In short, there are many conclusions we can and should draw from Russell's scenario.

In order to proceed with clarity, let us simplify Russell's argument. For now, we completely put aside the important question of whether the 'contained in' relation $\in$ is defined for all sets or not. We then consider the expression:

(0) $\langle \forall x :: \; x \in S \;\equiv\; \neg(x \in x) \;\rangle$,

where x is of type set. Thus we may reason as follows:

(0)

$\Rightarrow$ { instantiation, assuming S is a set }

$S \in S \;\equiv\; \neg(S \in S)$

$\equiv$ { predicate calculus }

false .

We therefore conclude:

$$\langle \forall S :: \ (S \text{ is a set}) \ \Rightarrow \ \neg(0) \ \rangle \qquad ,$$

or in English, "No set S satisfies (0) ." .

This conclusion is not new. In fact, it is the conclusion that Frege, Cantor, and Russell came to. So why is Russell's paradox still touted as a paradox? Presumably because Russell's scenario contains the hidden linguistic assumption that S is a set and satisfies (0) , an assumption that leads to a contradiction, as we have seen.

There is no paradox, simply the conclusion that a mathematical object satisfying (0) is not a set. (Assuming the relation $\in$ is defined for all sets.) Thus we can conclude that Frege and Cantor's proposed definition of a set is untenable.

*　　*
*

We should now return with interest to the question of whether $x \in y$ has a truth value for all x and y . To many, this question seems absurd, because they are used to thinking of 'truth' and 'falsity' as being some sort of real philosophical entities. (I cringe at the thought of whatever that means.) But since the time of George Boole, we can take a more sober view of things. The expression $x \in y$ is simply a mathematical expression, which can take on the values **true** and **false** , depending on the values of x and y . (Just like the expression $x + y$, for example.)

We have long been used to the fact that some mathematical expressions are "undefined" , or more accurately, "undefinable" , or even more accurately, "useless" . For example, in the language of arithmetic, the expression $0/0$ can be formed, but we can prove that $0/0$ is not a complex number, and hence it cannot be manipulated according to any of the familiar arithmetic rules. There is no contradiction here, just a useless expression, as useless as the expression $+14 + + = 3$.

In modern parlance, Frege and Cantor's mathematical concept would be called a 'class' , which is a generalization of the concept 'set' . There is a 'contained in' relation $\in$ for classes, but it is not defined for all classes.

Russell's S is a class; in particular, it is the class of all classes that do not contain themselves. Or in crisper, mathematical terminology, it satisfies:

$$(1) \qquad c \in S \ \equiv \ \neg(c \in c) \qquad ,$$

wherever this expression is defined. Following Russell's argument, we are led to the conclusion that $S \in S$ is undefined, and thus the question of whether S contains itself is simply unanswerable.

Like $0/0$, the expression $S \in S$ is useless. But there is no contradiction, and S is indeed a welcome creature in the world of classes. We can even prove things about it, for instance:

$$\begin{array}{ll} & \emptyset \in S \\ \equiv & \{ \ (1)\ ,\ \text{since}\ \emptyset\ \text{is a class}\ \} \\ & \neg(\emptyset \in \emptyset) \\ \equiv & \{ \ \text{definition of}\ \emptyset\ ,\ \text{the empty class}\ \} \\ & \neg(\mathbf{false}) \\ \equiv & \{ \ \text{predicate calculus}\ \} \\ & \mathbf{true} \qquad . \end{array}$$

Hence we have $\emptyset \in S$, or in English: "The empty set is in S." . Thus, some classes are in S , and, perhaps, some are not. And for yet others, it simply cannot be determined one way or the other.

The 'un-self-descriptive' paradox

I have little to add here, but I include this section because I find the "paradox" charming. The scenario is as follows: "The word 'pentasyllabic' is called self-descriptive, because it has five syllables, while the word 'tree' is called un-self-descriptive, because it is not a tree." .

What about the word 'un-self-descriptive' ? If 'un-self-descriptive' describes itself, then by definition it doesn't, and if it doesn't describe itself, then by definition it does. Mathematically modelling the situation, let w denote the word 'un-self-descriptive' , and let D denote the relation 'describes' . Then we have:

(0) $$w\,D\,x \ \equiv\ \neg(x\,D\,x) \qquad ,$$

for all words x . Thus we have recreated Russell's paradox, albeit in a different domain. Because w is a word, we may instantiate x in (0) with w , yielding **false** . Therefore, to the extent that we agree with property (0) , we are forced to conclude that D is not defined for all words.

Should we be surprised? We have already seen that even mathematical concepts like class membership may not always be defined, so it seems perfectly sensible that a fuzzy notion like 'description' can fail to be defined as well. This in no way prevents us from using the word! I can happily declare that the word 'tree' is un-self-descriptive, while 'pentasyllabic' is not. And as for the question of whether 'un-self-descriptive' is un-self-descriptive? Well, that question simply does not have an answer.

So once again, there is no paradox. The paradoxical feeling comes only from the hidden assumption that all natural language descriptions are well-defined. (Or put another way, that all yes-or-no questions have answers.)

JAW94–5

Self-referentiality

As we have seen, the nature of the above paradoxes has nothing to do with self-referentiality. The claim that they do, however, is not new. Douglas Hofstadter, for example, was clearly of this mind when he wrote his rich and diffuse popular science book *Gödel, Escher, Bach* . I would imagine that the late Willard Van Orman Quine held this view as well.

The most notable proponents of this position were presumably Bertrand Russell and Alfred Whitehead, the co-authors of *Principia Mathematica* . They constructed a strict hierarchical system of types, in an effort to keep self-referential paradoxes out of their version of mathematics. In this system, statements of a given type may refer to statements of a lower type only, thus effectively eliminating the possibility of self-reference. However, their typically philosophical —sorry, I couldn't resist— efforts were in vain, as Gödel's First Incompleteness Theorem would later demonstrate.

I should add that as mathematicians, we ought to be careful with sloppy word use. It seems clear to me a great deal of poetic license is needed in order to label an expression like $S \in S$ "self-referential" . That being said, it is also clear that expressions like these provide the means *par excellence* of creating contradictions. And such constructions are not without interest, as Gödel's Theorem shows.

References

Well, that concludes my little letter. I hope you enjoyed reading it. I'll close with a few references.

The discussion of the barber paradox is along the lines of the exposition in *Where is Russell's "Paradox"?* , a short position paper by the late Edsger W. Dijkstra. However, the formulation here is my own.

For the crisp calculational presentation of Russell's argument, I am indebted to another of Dijkstra's notes, *For brevity's sake* . In the interest of historical completeness, it should be added that Cantor claimed to have discovered Russell's paradox years earlier, but kept it a secret. (Little good it did him, the poor bastard.)

Santa Cruz, 1 June 2007

Jeremy Weissmann
jeremy@mathmeth.com