THE FALLACY IN RUSSELL'S SCHEMA

Hartley Slater
Philosophy / U. of Western Australia
Crawley, WA 6009, Australia
slaterbh@cyllene.uwa.edu.au

An analysis of the paradoxes of self-reference, which Bertrand Russell initiated, exposes the common fallacy in them, and has consequences for some of Graham Priest's work. Notably it undermines his defence of the Domain Principle, and his consequent belief that there are true contradictions. Use of Hilbert's epsilon calculus shows, instead, that we must allow for indeterminacy of sense in connection with paradoxes of self-reference.

Graham Priest has done a great service to the logical community by showing that most of the paradoxes of self-reference have a common structure. It is a structure which Bertrand Russell anticipated, and discussed in connection with some of the central paradoxes of self-reference.[1] But Priest has shown that a development of "Russell's Schema", which he calls the Inclosure Schema, fits almost all of the paradoxes.[2] As we shall see, however, there is a clear fallacy in Russell's Schema, which therefore enables us to escape not only Russell's Paradox, but also all the paradoxes of self-reference which have a sufficiently similar form.

Priest has also proposed that, since the paradoxes in question have a

[1] B. Russell, "On Some Difficulties in the Theory of Transfinite Numbers and Order Types" (1905), in D. Lackey, ed., *Essays in Analysis* (London: Allen and Unwin, 1973), p. 142.

[2] G. G. Priest, "The Structure of the Paradoxes of Self-Reference", *Mind*, 103 (1994): 25–34; Priest, *Beyond the Limits of Thought* (Cambridge: Cambridge U.P., 1995; 2nd ed., Oxford: Clarendon P., 2002), Pt. 3, Chap. 9, "Vicious Circles". My references are to the first edition.

n.s. 22 (winter 2002–03): 143–50
ISSN 0036-01631

uniform structure, they ought to have a uniform solution, and, since he thinks the Inclosure Schema is sound, he has come to believe that *Reductio* should be discarded. The uniform solution of the paradoxes is given below, however, by identifying the common fallacy. The reason *Reductio* is not in danger is that Priest's own analysis straightforwardly shows there aren't the inescapable contradictions many have thought. If a contradiction is derivable from some assumptions, the conjunction of them is shown to be false; but Priest has been so persuaded of certain assumptions he has been led, instead, to affirm *as true* the contradiction derivable from them. I shall show there is no doubt that one of the assumptions Priest believes to be true is false.

Priest first shows that several of the paradoxes of self-reference follow exactly Russell's Schema:

> … given a property ϕ, and function δ, consider the following conditions:
>
> 1) $w = \{x : \phi(x)\}$ exists,
> 2) if x is a subset of w: a) $\delta(x) \notin x$,
> and b) $\delta(x) \in w$
>
> … Given (1) and (2) we have a contradiction. For when [they] are applied to w, an irresistible force meets an immovable object.[3]

He then illustrates the matter in a central case:

> In Russell's Paradox, the property $\phi(x)$ is "$x \notin x$", so that w is the Russell set $R = \{y : y \notin y\}$; and the function δ is simply the identity function, id. Suppose that $x \subseteq w$; then $x \in x \Rightarrow x \notin x$. Hence $x \notin x$ …; it follows that $x \in R$.… The contradiction is that $R \in R$ and $R \notin R$.[4]

But Russell did not distinguish

$$(\exists y)(y = \varepsilon x(z)(z \in x \equiv z \notin z)),$$

and

$$(\exists x)(z)(z \in x \equiv z \notin z).$$

[3] Priest's paraphrase in *Mind*, p. 27; *cf. Beyond the Limits*, p. 142. As Priest notes, he has "taken the liberty of modifying [Russell's] notation" (p. 142 n.4).

[4] Priest, *Mind*, p. 27; *cf. Beyond the Limits*, p. 142.

The first says the Russell Set exists ($R = \varepsilon x(z)(z \,\varepsilon\, x \equiv z \not\varepsilon\, z)$), the second says there is such a set, i.e. a set having the supposed properties of the Russell Set ($(z)(z \,\varepsilon\, x \equiv z \not\varepsilon\, z)$). The first is necessarily true, but the second is false, indeed contradictory, which means there are no grounds for the deduction of "$R \,\varepsilon\, R$" and "$R \not\varepsilon\, R$" as truths.

More specifically, if $x \subseteq w$ in this case, then while $x \,\varepsilon\, x$ consequently entails $x \,\varepsilon\, w$, one cannot get from this $x \not\varepsilon\, x$, since that requires the second existence statement to be true, and only the first is.

The epsilon calculus in which these distinctions can be made was formulated by David Hilbert in the 1920s. Epsilon terms are defined for all predicates in the language, which makes the first existence statement true. But in epsilon calculi a definition of the existential quantifier can be formulated, viz, "$(\exists x)Fx \equiv F \varepsilon xFx$", with the result that the above epsilon term would have to live up to its name, i.e be *attributive* for the second existence statement to be true. Since that statement is false, however, the term must be *non-attributive*, which means it does not define any set. But that does not mean it defines, say, a proper class instead; it merely means its reference is arbitrary.[5] If the epsilon term in the first existence statement was a Russellian iota term, then the two existence statements would be equivalent, but a discrimination is available using the epsilon term. This is because, unlike an iota term, an epsilon term is a complete term for an individual, i.e. what Russell called a logically proper name. Frege, remarkably, used a complete term for "the F" in some of his formal work, but Russell, of course, argued against the arbitrariness of Frege's definition in the case where there is not just one F, when setting up his alternative theory of descriptions in "On Denoting". It is just this feature of logically proper names, however, which gets us out of the paradoxes which Russell did so much to dramatize.

In his treatment of other paradoxes Priest adds a further condition to the clauses above, to produce the Inclosure Scheme,[6] but that does not affect the above point; hence there is no inescapable contradiction in any case. Priest traces his defence of the (second form of the) existence clause back to what Michael Hallett called the Domain Principle.[7] Cantor

[5] B. H. Slater, "Hilbertian Reference", *Nous*, 22 (1988): 283–97.

[6] Priest, *Mind*, p. 28; *Beyond the Limits*, p. 147.

[7] M. Hallett, *Cantorian Set Theory and Limitation of Size* (Oxford: Clarendon P., 1984), p. 7; Priest, *Beyond the Limits*, p. 137.

used the Domain Principle to obtain a completed infinity from every potential infinity, and Priest uses it to defend set-theoretical semantics even to the extent of requiring the naive Abstraction Axiom to be true. But the Domain Principle is predicated on a certain assumption, which Priest is quite explicit about, to do with determinacy of sense (*Beyond the Limits*, p. 139), and this assumption is just what is violated in the paradoxical cases in question. Maybe we can start to formulate the notion of Heterologicality, for instance, from facts like: "short" is short, "long" is not long. But we cannot complete the process, since we would not know what to say about "heterological" itself. In the case of Richard's Paradox, given an enumeration of all real numbers definable in a finite number of words, a further real number, not in that list, can be defined. But what that shows is that there isn't a clear list of all of them: "As pointed out by [Borel], the process of throwing away those symbol strings which do not define real numbers cannot be carried out effectively."[8]

Not only did Priest insist there still was a determinate set in such cases, however, he thought he therefore could abandon the procedure we have for discriminating the cases where there isn't a determinate set—upon the appearance of a contradiction, using *Reductio*—which obviously left him without a criterion to eliminate the confusing cases. But it also, more crucially, led him away from seeing what replaces sets. Priest wanted to say that Ramsey's distinction between set-theoretical and semantical paradoxes was ungrounded—because all were in fact set-theoretical.[9] But the parallelism instead runs the other way: the semantic definability paradoxes, like Heterologicality, and Richard's, are the paradigms, and show that some sets are not definable—a conclusion which extends also to the set-theoretic paradoxes, like Russell's above.

In fact, that there is no such set as the Russell Set is no different in kind from, for example, there being no natural number between 2 and 3. For, although we can suppose there is a natural number between 2 and 3, i.e. say

$$(\exists n)(n > 2 \,.\, n < 3),$$

[8] P. Martin-Löf, *Notes on Constructive Mathematics* (Stockholm: Almqvist and Wiksell, 1970), p. 44.

[9] Priest, *Mind*; *Beyond the Limits*, Pt. 3, Chap. 10.

and even go on to talk about the number in question, i.e.

$$\varepsilon\, n(n > 2 \,.\, n < 3) > 2 \,.\, \varepsilon\, n(n > 2 \,.\, n < 3) < 3,$$

using the epsilon axiom

$$(\exists n)Fn \supset F\,\varepsilon\, nFn,$$

further analysis shows that this leads to a contradiction. We find we must have

$$\neg(\varepsilon\, n(n > 2 \,.\, n < 3) > 2 \,.\, \varepsilon\, n(n > 2 \,.\, n < 3) < 3),$$

which is just to say

$$\neg(\exists n)(n > 2 \,.\, n < 3).$$

So although we can form a referring phrase in this case, it simply turns out to be non-attributive, a fact we commonly put by saying that there is no "such" number as the number in question. The referential phrase "$\varepsilon\, n(n > 2 \,.\, n < 3)$" still has an arbitrary reference, by the definition of epsilon terms,[10] i.e.

$$(\exists m)(m = \varepsilon\, n(n > 2 \,.\, n < 3)),$$

but there is no natural number with the character, or description "$n > 2 \,.\, n < 3$", i.e.

$$\neg(\exists n)(n > 2 \,.\, n < 3).$$

Priest has appealed to the modern set-theoretic analysis of quantification in defence of the Domain Principle, and he covered alternatives to sets in this respect in the later chapters of his book, discussing there the theory of proper classes, and the use of plurals. But he should have

[10] A. C. Leisenring, *Mathematical Logic and Hilbert's* ε*-Symbol* (London: Macdonald, 1969).

looked more closely at the definition of the quantifiers in the epsilon calculus, as we can now see. Priest has not ignored the epsilon calculus, indeed he has explored its relevance to Berkeley's Paradox at length. But Priest there explicitly objected to moves like that from

$$\neg(\varepsilon\, n(n > 2 \,.\, n < 3) > 2 \,.\, \varepsilon\, n(n > 2 \,.\, n < 3) < 3),$$

to

$$\neg(\exists n)(n > 2 \,.\, n < 3),$$

since he claimed that the epsilon axiom does not contrapose (*Beyond the Limits*, pp. 73–4). But the epsilon definition of the quantifiers, i.e.

$$(\exists n)Fn \equiv F\,\varepsilon\, nFn,$$
$$(n)Fn \equiv F\,\varepsilon\, n\neg Fn,$$

which he also has to drop as a result, is just what avoids reference to determinate sets in the way Priest thinks the Domain Principle requires. A set-theoretic semantics for certain epsilon calculi can be given (Leisenring, pp. 19, 40), but that requires acceptance of a second epsilon axiom:

$$(x)(Fx \equiv Gx) \supset \varepsilon\, xFx = \varepsilon\, xGx.$$

And Hilbert, in fact, introduced epsilon terms expressly to eliminate the quantifiers.

Priest discusses a telling case from Patrick Grim. Grim wanted to say that there was no set of all propositions, but realized this needed some clarification: "But the denial that there is any such thing as '… all propositions' should not itself be thought to commit us to quantifying over … all propositions, any more than the denial that there is such a thing as 'the square circle' should be thought to commit us to referring to something as both square and a circle."[11] Indeed, $\neg(\exists x)(Sx \,.\, Rx)$ still allows $(\exists y)(y = \varepsilon\, x(Sx \,.\, Rx))$; it only means that $\neg(S\,\varepsilon\, x(Sx \,.\, Rx) \,.\, R\,\varepsilon\, x(Sx \,.\, Rx))$, i.e. that the associated epsilon term is non-attributive. How can Grim say "there is no totality of all the propositions"? Priest

[11] P. Grim, *The Incomplete Universe: Totality, Knowledge and Truth* (Cambridge, MA: MIT P., 1991), p. 123.

thinks that this thesis is inexpressible, on its own account, since it quantifies over all propositions, and so a totality of them is presupposed (*Beyond the Limits*, p. 253). But if ϕy is "y is a proposition", then Grim's remark is

$$\neg(\exists x)(y)(\phi y \equiv y \,\varepsilon\, x),$$

which simply becomes

$$\neg(\phi a \equiv a \,\varepsilon\, b),$$

for certain constants a and b, on the epsilon reduction.

What confuses is not just that we can start on the job of generating a set in the paradoxical cases which we necessarily cannot finish; also, and even more significant, we allow, in the process, reference to things which only certainly exist in name only—the sets which are then in prospect. We can discriminate propositions, for instance, and so come to envisage the totality of them; but, as Grim shows, there is no such totality (p. 122). So talking about that totality does not guarantee it exists *as a totality*, i.e. with its supposed character. "There is no such set as the Russell Set, i.e. that set defined in connection with Russell's Paradox", we may say. But that does not mean the Russell Set exists in more than a nominal sense: "that set" refers, but it does so non-attributively. Anyone taken in by the existence clause in Russell's Schema is, by contrast, thinking that being able to suppose there is a set of a certain kind means there is in fact such a set. Without a presumption of attributivity there would be no feeling of paradox upon finding that, in certain central cases, there simply is nothing of the kind in question. We escape the paradoxes not through any technical adjustment, therefore, but simply through coming to see that certain contradictory results are quite natural, on the assumption of attributivity. The surprise, if it is there, lies is finding that one has been dealing with a logically proper name for an individual all along. But one was simply in error to expect otherwise.

So also: it is simply *Reductio* which rules out the attributivity of the referential term in question, when it is applicable. That is the way we know there is no such concept as Heterologicality—just the way we know there is no barber who is such that he shaves all and only those who do not shave themselves. Why is there no set whose members are all and only those sets which are not members of themselves? Because of Russell's contradiction. Expecting all terms of the form "$\{x|Fx\}$" to be attributive was what led to Frege's downfall—the double irony being that Frege elsewhere formulated individual terms which were non-attributive, while Russell argued that we must read all definite descriptions—which therefore had to include terms like "the set of things which are F"—as attributive. But to expect there to be some other way of specifying which terms of the form "$\{x|Fx\}$" are non-attributive than by using *Reductio* is equally fallacious. It is like expecting there to be some other way to determine which terms like "the natural number which is F" do not describe numbers. Priest's thinking we can do away with *Reductio* is thus just as illogical.