Reviews

Russell's correspondence with Frege

by David Bell

Gottlob Frege. *Philosophical and Mathematical Correspondence*. Abridged from the German edition by B. McGuinness. Translated by H. Kaal. Oxford: Basil Blackwell; Chicago: University of Chicago Press, 1980. Pp. xviii+214.* £17.50. US$31.00.

1. *Introduction*

Russell was just twenty-eight, and had been intensively involved in the study of mathematical logic for less than a year, when he discovered that certain intuitively plausible and apparently harmless assumptions, widely made by logicians, are in fact provably incoherent: they lead to the contradiction which has since come to be known as Russell's paradox. In its most virulent form the paradox concerns the notion of a class and, in particular, the notion of a class of classes. Now, if classes are logically respectable entities, then there seems no good reason why we should not allow that there can be classes *of* such entities. If classes of classes are allowed then, trivially, a class may be a member of a class. But in order that the notion of a class should not remain dangerously indeterminate, it must be decided, one way or the other, whether a class may be a member of itself. Now intuitively at least, this much seems clear: there are some classes which are not members of themselves. The class of elephants, for example, is not itself an elephant and is not, therefore, itself a member of the class of elephants. And again, if the notion of a class which is not a member of itself is logically respectable, there seems no good reason why we should not talk about the class of all such classes (we can call this class "*C*" for short). We must have gone wrong somewhere, however, because

* Originally published in German as *Wissenschaftlicher Briefwechsel*, ed. G. Gabriel, H. Hermes, F. Kambartel, C. Thiel, and A. Veraart (Hamburg: Felix Meiner Verlag, 1976). Pp. xxviii+310.

class *C* gives rise to the following absurdity: if *C* is a member of itself then it is not a member of itself while, conversely, if *C* is not a member of itself then it is a member of itself. This is the contradiction which Russell communicated to Frege, one year after its discovery, in his first letter, June 16th, 1902.

Sixty years later Russell described the reaction of his correspondent in the following terms:

> As I think about acts of integrity and grace, I realise that there is nothing in my knowledge to compare with Frege's dedication to truth. His entire life's work was on the verge of completion, much of his work had been ignored to the benefit of men infinitely less capable, his second volume [i.e. of the *Basic Laws of Arithmetic*] was about to be published, and upon finding that his fundamental assumption was in error, he responded with intellectual pleasure clearly submerging any feelings of personal disappointment. It was almost superhuman....[1]

While there is some truth in the portrait Russell presents here, the sentimental, even fulsome tone is quite misplaced and serves only to mask a number of distortions. It seems in fact likely that Russell never fully realized the terrible effect his first, short, almost chatty letter to Frege had on its recipient. Frege's reaction was not "superhuman", and one can be sure that he took very little "intellectual pleasure" in the information which Russell communicated to him. On the contrary, Frege quickly abandoned completion of what was to have been his magnum opus, *The Basic Laws of Arithemtic*: he produced virtually no work of any sort for the next seven years; and he lived his last years as a broken, disillusioned, and bitter man. When Russell discovered the contradiction he was twenty-eight and had been immersed in mathematical logic for less than a year; when Frege learned of it he was almost twice Russell's age, and had single-mindedly devoted his intellectual efforts to the development of mathematical logic for over twenty years. As Frege himself wrote, with admirable constraint: "Hardly anything more unwelcome can befall a scientific writer than that one of the foundations of his edifice should be shaken, after the work is finished. I have been put in this position by a letter from Mr. Bertrand Russell."[2]

[1] Letter from Russell to van Heijenoort, November 23rd, 1962. Published in J. van Heijenoort, ed., *From Frege to Gödel* (Cambridge, Mass.: Harvard Univ. Press, 1967), p. 127.

[2] Frege, *The Basic Laws of Arithmetic*, ed. and trans. M. Furth (Berkeley: Univ. of California Press, 1964), p. 127.

The contradiction posed so very great a threat to both Russell and Frege for the same reason: both had, independently, identified cardinal numbers with classes of classes. Frege replied to Russell's letter very quickly, virtually by return of post, acknowledging that the latter's discovery "seems to undermine not only the foundations of my arithmetic, but the only possible foundations of arithmetic as such" (p. 132). Although Frege spent some time investigating means by which the contradiction could be avoided, he eventually came to believe that Russell *had* undermined the only possible logical foundation of arithmetic. As he wrote in his diary shortly before his death: "My efforts to become clear about what is meant by number have resulted in failure."[3] Frege's reasons for this melancholy conclusion, and in particular his reasons for rejecting all Russell's proferred solutions to the paradox, emerge clearly in the course of their correspondence—and it is on this topic that I shall concentrate in what follows. This means that I shall ignore those frequent and often protracted passages in which Frege attempts—though to little avail—either to correct Russell's misunderstandings about the *Basic Laws* or to get him to be logically more rigorous. In this latter respect Gödel's verdict is incontestable: "It is to be regretted", he wrote about *Principia Mathematica*, "that it is so greatly lacking in formal precision in the foundations that it represents in this respect a considerable step backward as compared with Frege. What is missing, above all, is a precise statement of the syntax of the formalism. Syntactical considerations are omitted even when they are necessary for the cogency of proofs, in particular in connection with the 'incomplete symbols.'"[4] Russell should have learned more than he did from Frege, but I shall not dwell on the lessons that remained unlearned.

2. *The contradiction communicated*

This is how Frege first came to learn of the contradiction. Russell wrote:

> I have encountered a difficulty only on one point. You assert ([*Begriffsschrift*] p. 17) that a function can also constitute the indefinite element. This is what I used to believe, but this view now seems to me to be dubious because of the following contradiction. Let *w* be the predicate: to be a predicate that cannot

[3] Frege, *Posthumous Writings*, ed. H. Hermes *et al.*, trans. P. Long and R. White (Oxford: Blackwell, 1979), p. 263.

[4] K. Gödel, "Russell's Mathematical Logic" in *The Philosophy of Bertrand Russell*, ed. P.A. Schilpp, 4th ed. (La Salle, Ill.: Open Court, 1971), p. 126.

> be predicated of itself. Can w be predicated of itself? From either answer follows its contradictory. We must therefore conclude that w is no predicate. Likewise there is no class (as a whole) of those classes which, as wholes, are not members of themselves. From this I conclude that under certain circumstances a definable set does not form a whole. (P. 130)

Apart from a brief postscript which expresses the contradiction in Peano's notation, this is all the discussion the topic receives in Russell's first letter, and I have quoted it in full in order to counter a number of widespread but mistaken beliefs about the extent at this time of Russell's understanding of the contradiction and of how it arises within Frege's system. Quine, for example, says: "Russell wrote Frege announcing Russell's paradox and showing that it could be proved in Frege's system."[5] And van Heijenoort has claimed that Russell correctly identifies the passage in the *Begriffsschrift* (i.e. p. 17) which is responsible for the "flaw in Frege's system".[6] In fact, however, Russell demonstrates neither that, nor where, nor how Frege's work is inconsistent. The passage in the *Begriffsschrift* to which he alludes, and which indeed mentions "indeterminate elements" in a judgment, is not at all what gives rise to the contradiction. On the contrary, and ironically, it is at precisely this point in the *Begriffsschrift* that Frege introduces, for the first time in the history of logic, the notion of a second-level function whose arguments are always first-level functions and whose values are either true or false—the notion, in other words, of a quantifier. It is precisely this syntactic distinction between proper names, first-level function-names, second-level function-names, and so on which *prevents* Russell's paradox of self-predication from infecting Frege's logic. As Frege himself says in his first letter to Russell: "the expression 'a predicate is predicated of itself' does not seem exact to me. A predicate is as a rule a first-level function which requires an object as argument and which cannot therefore have itself as argument" (p. 132). Frege's logical syntax ruled that expressions of the form: "$\phi(\phi)$" are simply malformed. And because the *Begriffsschrift* employs no set-theoretic apparatus the paradox cannot be derived within it. (Van Heijenoort is further mistaken in that he takes Russell to be alluding to a quite different, but equally harmless, doctrine which appears on page 19 of the *Begriffsschrift*.)

3. *Russell's first suggestion*

Russell's next letter (XV/3) contains his first, tentative suggestion as to how the contradiction might be avoided. "The contradiction only arises if the argument is a function of the function," he writes, "that is, only if argument and function cannot vary independently" (p. 133). He cites two kinds of case in which this may happen: first, the case in which a concept is predicated of its own extension, *viz.* "$\phi(\hat{x}(\phi x))$";[7] and secondly the case in which a concept is predicated of itself: "$\phi(\phi)$". The suggestion implicit in Russell's letter is that if both forms of expression were outlawed there would be no possibility of generating the contradictions. Frege replies (XV/4) that the cost of rejecting all expressions of the first sort is too high; for the law of excluded middle would also therewith have to be abandonned. If "ϕ" is a significant predicate, and if "$\hat{x}(\phi x)$" is the name of an object, then either "$\phi(\hat{x}(\phi x))$" or its negation must be true, *tertium non datur*. The only way to avoid such exceptions to the law of excluded middle, Frege argues, would be to deny that, properly speaking, classes are *objects* at all. But for Frege an object is that which can be the reference of a singular term, and a singular term has reference just in case it can participate in direct discourse which possesses a determinate truth-value. That classes are proper objects is thus for Frege an immediate consequence of the fact that names like "$\hat{x}(\phi x)$" can occur in sentences possessed of a determinate truth-value—for example, trivially, in "$\hat{x}(\phi x) = \hat{x}(\phi x)$". As to Russell's suggestion that one avoid expressions of the second, self-predicative sort, Frege has already pointed out that a correct logical synatax will not allow them in the first place.

4. *Russell's second suggestion*

Inauspiciously, Russell begins his next letter to Frege by saying: "Concerning the contradiction I did not express myself clearly enough. I believe that classes cannot always be admitted as proper names [*sic*]. A class consisting of more than one object is, in the first place, not *one* object but many" (p. 137). Despite his chronic failure to distinguish between names and their bearers, and despite the obscurity of his contrast between classes as one and classes as many, the gist of Russell's second suggested solution to the paradox is clear enough. If we restrict consideration to classes with more than one member then, Russell claims, all classes are necessarily classes *as many*, whereas not every such class is also a class *as one*, i.e. a unitary whole whose members comprise its parts. The contradiction, it is claimed, arises as a result of taking a

[5] W. V. Quine, "On Frege's Way Out", *Mind*, 64 (1955): 147.

[6] Van Heijenoort, *From Frege to Gödel*, p. 3.

[7] At this time, unlike Frege, Russell had no notation for classes. Here I shall use Russell's later notational device: "$\hat{x}(\phi x)$" is the name of *the class of all ϕs*. What differences there are between sets, classes, extensions of concepts, and ranges of values of concepts are not here relevant, and I have used these terms interchangeably—as do Russell and Frege in their correspondence.

class as many (a manifold) to be a class as one in cases where this move is in fact unwarranted. There is thus no need to prohibit *all* expressions of the form "$\phi(\hat{x}(\phi x))$", but only those which contain a name "$\hat{x}(\phi x)$" which names a manifold but does not name a class as one.

Frege's reply is masterful. He begins by pointing out that classes as one and classes as many (manifolds) cannot be said to differ in that one is while the other is not an object. In so far as we use a class name like "$\hat{x}(\phi x)$" in sentences with a determinate truth-value we treat a class as a single, self-identical object. And in so far as we use the same sign—or even a different sign, so long as it is a singular term—to talk about a manifold, we are treating the manifold too as a single, self-identical object.

Frege then points out that in so far as the language of parts and wholes is appropriate at all in this context, it applies only to the relation between a manifold and its elements, and never to the relation between a class and its members, contrary to what Russell believed. The reasons for this emerge as Frege proceedes to specify precisely the differences between a class and a manifold, and to show why the latter cannot possibly be used in a logical definition of number. (In what follows I shall use the term "whole" to mean a complex comprising a number of elements or parts, i.e. a manifold.) Frege identifies six crucial differences between wholes and classes:

(i) A whole is a complex whose parts are united by a system of relations one with another. Hence the whole can be destroyed by dissolving the relations, even if the parts survive the dissolution: an army can be destroyed by being disbanded, even though all the individual soldiers which comprise it survive; a forest can be destroyed by dispersing the trees, even though each of these continues to exist elsewhere. The same is not true of a class, which is not "held together" by the relations of its members to one another. It cannot therefore be destroyed if these relations change.

(ii) A whole depends for its existence on the existence of its parts: destroy the parts and the whole is thereby destroyed. The same is not true of a class and its members; for "$\hat{x}(\phi x)$" is still the name of a class even if there are no ϕs. In this case it is the name of the null class.

(iii) A whole does not have a unique decomposition into parts. There is, in other words, no determinate answer to the question "Which things are the parts of this whole?" Thus an army can be said to be composed of regiments, or batallions, or companies, or soldiers ... and so on. By contrast, as Frege says, "when we are given a class it is determined what objects are members of it" (p. 140).

(iv) Likewise a whole does not have a determinate number of parts, whereas a class does have a determinate number of members.

(v) The relation of a whole to its parts is governed by the principle that a part of a part is also a part of the whole. There can be no analogous principle governing the relation of a class to its members: the relation of membership is not transitive. (This explains the asymmetries noted in (iii) and (iv) above.)

(vi) A whole whose parts are material things is itself a material thing. A class, however, is always a logical or abstract object—even when its members are material things. The relation of a whole to its parts is thus a material, not a logical relation.

The asymmetry noted in (iv) makes the whole/part relation unsuitable as a basis on which to develop a theory of *number*; and the asymmetry noted in (vi) makes it unsuitable as a basis on which to develop a purely *logical* account of number. Point (ii) also presents severe difficulties for anyone wishing, say, to identify the number zero with the class whose sole member is the null class. For if a class is taken to be *composed of* its members the very notion of a "null class" becomes problematic.

Russell appears to have been convinced by Frege's objections, for his next letter begins: "Many thanks for your explanations concerning ranges of values [i.e. classes]. I now understand the necessity of treating them not merely as aggregates of objects" (p. 143). The understanding did not last long. Appendix B of *The Principles of Mathematics* was written some three months after this letter, and there we find Russell again reverting to the mistaken assimilation of classes to concrete wholes: "the objects of daily life, persons, tables, chairs, apples, etc. are classes as one."[8]

5. *Russell's third suggestion*

Russell's letter of August 8th, 1902 (XV/9) is notable in that it contains one of the very earliest formulations of a *theory of logical types* for classes. Interestingly, Russell introduces the suggestion explicitly as an extension of Frege's distinction between objects, first-level functions, second-level functions and so on:

> The contradiction could be resolved with the help of the assumption that ranges of values are not objects of the ordinary kind; i.e., that ϕx needs to be completed ... either by an object, or by a range of values of objects, or by range of values of ranges of values, etc. This theory is analogous to your theory about functions of the first, second etc. levels. (P. 144)

For every function ϕx there would accordingly be not only a range of values

[8] *The Principles of Mathematics*, 2nd ed. (London: Allen & Unwin, 1937), p. 522.

but also a range of those values for which ϕx is decidable, or for which it has a sense. (P. 145)

Frege reacted to this suggestion without enthusiasm, pointing out that the law of excluded middle will still be threatened unless (a) precise *syntactic* rules can be formulated which not only perspicuously determine the well-formedness or otherwise of every possible function name when its argument-place is filled by an expression of type-0, type-1, type-2, and so on; and unless (b) a *value* is specified for every function for every level of class as permissible argument. Frege believed that this could not in practice be done. But even were it in theory possible, he believed the suggestion was also objectionable because it would threaten the *generality* of arithmetical truths, and their fundamentally *logical* nature. Not only would extraneous and largely ad hoc devices be introduced into logic, but also the numbers would need to be defined afresh at each level in the type hierarchy.

6. *A new contradiction*

"My proposal concerning logical types now seems to me incapable of doing what I had hoped it would do", Russell wrote (XV/11) a few days after receiving Frege's objections. It is clear, however, that it was not the latter that caused Russell's change of mind. Rather, he had discovered, he believed, yet another contradiction—one moreover whose derivation even adoption of a theory of types for classes would fail to block. The new antinomy is not blocked because it concerns, not classes of classes, but classes of propositions.

The text of Russell's letter is dense and extremely obscure, and although some light is shed on it by §500 of the *Principles* where the same contradiction is also mentioned, this section too is very far from clear. For this reason, then, and also because a number of subsequent letters are devoted to the new antinomy, it will be worthwhile attempting to spell out what Russell had in mind. This is how Russell expressed the problem to Frege:

> If m is a class of propositions, then "$p \varepsilon m \cdot\supset_p\cdot p$" represents their logical product. This proposition itself can either be a member of class m or not. Let w be the class of all propositions of the above form which are not members of the pertinent class m, i.e.,
>
> $$w = p\ni(\exists m\ni\{p \;.=:\; q\varepsilon m \;.\supset_q.\; q:.p \sim\varepsilon m\});$$
>
> and let r be the proposition $p\varepsilon w \cdot\supset_p\cdot p$.
>
> We then have $r\varepsilon w \cdot\equiv\cdot r \sim\varepsilon w$.
>
> Here we must consider the content of the propositions, not their meaning; and we must not take equivalent propositions to be simply identical. (P. 147)

Before we can begin to extract an argument from this it is necessary to clear away a couple of confusions on Russell's part which make it extremely difficult to establish exactly what he meant. In the first place, Russell subscribed at this time to a faulty account of the logical product of two or more propositions: the definition of "logical product" which he provides at §18 of the *Principles*, and which he subsequently employs in the discussion of the new antinomy at §500, simply does not work. This, however, is merely a technical slip; we can assume that Russell meant by "the logical product of p and q" what *we* now mean by the phrase, i.e. what today would be symbolized as "$p\&q$". In 1903, however, Russell did not employ any sign corresponding to "&", but represented the logical product of propositions by immediate concatenation: "pq". This notation no doubt encouraged him in the second, much more serious confusion of which he was guilty, namely identifying *the class* of propositions p and q with *the logical product* of those propositions: to someone who believed that a class was "composed of" its members, doubtless pq seemed very similar to $\{p, q\}$. In this way Russell confused a singular term, a class name, with a complex proposition—and this helps to explain the peculiar opening sentence of the above quoted passage. In fact it was precisely this confusion of a class of propositions with the logical product of those propositions which enabled Russell to apply Cantor's paradoxical results to the case in hand. This is surprising, for the antinomy bears a striking resemblance to the Paradox of the Liar and was, indeed, later handled by Russell in that guise.[9] It seems, however, that at this time Russell had yet to see any link between the modern, Cantorian mathematical contradictions and the ancient antinomies: there is, for example, no mention of the Liar either in the correspondence with Frege, or in the first edition of the *Principles.*

Putting Russell's confusions to one side, then, his argument would seem to be as follows. Let m be a class of true propositons, i.e. such that

$$a: \qquad p\varepsilon m \;.\supset_p.\; p$$

but, also, such that

$$b: \qquad \sim(a\varepsilon m).$$

[9] Russell, "Mathematical Logic as Based on the Theory of Types" in R. C. Marsh, ed., *Logic and Knowledge* (London: Allen & Unwin, 1956), pp. 59, 62, etc.

In other words all the propositions in class m are true, but proposition a is not itself a member of m. Next, let w be the class of all propositions of the same form as proposition a, such that for each such proposition the appropriate verison of b is true. So, for example, in addition to a, class w will also contain the following two propositions c and d:

$$c: \qquad p\varepsilon u \;.\supset_p.\; p$$

as long, that is, as $\sim(c\varepsilon u)$; and

$$d: \qquad p\varepsilon v \;.\supset_p.\; p$$

as long, that is, as $\sim(d\varepsilon v)$.

We can now examine a third candidate for membership of w, namely

$$r: \qquad p\varepsilon w \;.\supset_p.\; p$$

given, that is, that $\sim(r\varepsilon w)$. Unfortunately this implies that $r\varepsilon w \;.\equiv.\; \sim(r\varepsilon w)$.

The exact nature of this new antinomy and the threat it poses occasion much of the remaining correspondence between the two philosophers during the next two years. Somewhat surprisingly it is in fact *this* contradiction rather than the earlier and more famous one which prompts them to raise the most fundamental issues concerning, for example, the nature of sense, reference, truth, thought, functions, objects, proper names, direct and indirect discourse, propositions, and much else besides. In essence Frege's reaction was, steadfastly, that the new antinomy is merely apparent, and in fact results from Russell's failure to distinguish clearly the thought, the sense expressed by a sentence, from the truth-value (if any) which the sentence possesses. In effect his objection takes the form of a dilemma: either a proposition stands for its truth-value, in which case materially equivalent propositions can be treated as identical and, as Russell acknowledges, the contradiction does not arise; or, on the other hand, a proposition stands for a thought, in which case materially equivalent propositions will not in general be identical. But if a proposition stands for a thought, and in particular if the classes m, w, etc. mentioned in the derivation of the antinomy are taken to be classes of thoughts, then again the contradiction cannot be validly generated. For there will be an irreducible difference between a class of thoughts, on the one hand, and a thought about that class of thoughts on the other hand—just as, analogously, there will be a difference, say, between a class of elephants and a thought about a class of elephants. It is only Russell's blurring of this distinction which allows him to obtain his result; and this would appear to be a direct consequence of his confusing a class of propositions with the logical product of those propositions, noted earlier.

7. *Russell's fourth suggestion*

On May 19th, 1903 Russell had yet another idea as to how the contradictions might be avoided: "The relief of this is unspeakable" he wrote in his journal four days later; and five days later, on May 24th, he communicated his result to Frege. He had discovered, he believed, "that classes are entirely superfluous" (p. 158), and that it is perfectly possible "to do arithmetic without classes ... this seems to me to avoid the contradictions" (p. 159).

We can perhaps see here, in embryonic form, a line of thought which was eventually to culminate in the doctrine that classes "can be regarded as symbolically constructed fictions."[10] For the moment, however, the suggestion remained undeveloped, indeed barely coherent. Russell believed that one could quite simply dispense with a notation for classes and instead employ, in isolation, the corresponding function-names. He is therefore prepared to allow that an expression such as "$\phi \subset \psi$" is well formed.[11]

Frege did not reply to this letter until some eighteen months later. He blames "various distractions" for the delay, but it is likely that his interest in the matter was waning, and in fact this reply (XV/18) is the last occasion on which he wrote to Russell about philosophical matters. Presumably he was beginning to recognize that Russell's paradox could not be avoided by the restriction he had proposed in the Appendix to the second volume of the *Basic Laws*, or indeed by any other means which he was prepared to adopt, and that therefore his entire logicist programme would have to be abandoned. His last substantial letter, however, points out clearly the major failings of Russell's attempt to dispense with classes. It involves employing signs, such as "ϕ", in such a way that they "would be defined as a function sign and used as a proper name" (p. 162). Moreover, even if one were to adopt this procedure, the contradiction can still be generated, as Frege succinctly proves.

[10] Russell, "The Relation of Sense-Data to Physics", §VI, in *Mysticism and Logic* (London: Allen & Unwin, 1963), p. 115.

[11] In the English edition of the correspondence considerable confusion is created because the sign for material implication "$\supset$" has at times been substituted for the sign for class inclusion: "$\subset$" (e.g. p. 159, l. 4, and p. 161, l. 15). The German edition is more reliable.

8. *Conclusions*

The twenty letters which passed during this period of just over two years between Frege and Russell together comprise a rich source of historical information and philosophical stimulation—by no means every aspect of which has been touched upon in this review. In particular it is fascinating to witness the speed and inventiveness with which Russell was able to suggest radically new solutions to the contradictions. Indeed, we have in these letters some of the very earliest, tentative, and often naïve formulations of doctrines which were later to become central tenets of Russell's mature philosophy: the theory of types, for example, the doctrine that the meaning of a proper name is the object it refers to and hence that names do not express a sense, and also the "no-class theory". In this respect the collection of letters will be of greater use to those studying Russell than it will be to those studying Frege: throughout the correspondence Frege steadfastly continued to defend the doctrines which he had already published. Indeed it appears that as a result of their exchange of ideas Russell managed to change Frege's mind about only one major point. This was no small thing, however; for what Frege came to see was that his life's work was in ruins. The letters together provide us with a portrait of a genius who, though he was capable of winning virtually every battle, finally and tragically lost the war.

Department of Philosophy
Sheffield University