Russell’s Paradox

First published Fri Dec 8, 1995; substantive revision Mon Oct 12, 2020

Russell’s paradox is the most famous of the logical or set-theoretical paradoxes. Also known as the Russell-Zermelo paradox, the paradox arises within naïve set theory by considering the set of all sets that are not members of themselves. Such a set appears to be a member of itself if and only if it is not a member of itself. Hence the paradox.

Some sets, such as the set of all teacups, are not members of themselves. Other sets, such as the set of all non-teacups, are members of themselves. Call the set of all sets that are not members of themselves “$R$.” If $R$ is a member of itself, then by definition it must not be a member of itself. Similarly, if $R$ is not a member of itself, then by definition it must be a member of itself.

Although also noticed by Ernst Zermelo, the contradiction was not thought to be important until it was discovered independently by Bertrand Russell in the spring of 1901. Since then, the paradox has prompted a great deal of work in logic, set theory and the philosophy and foundations of mathematics.

1. The Paradox

Central to any theory of sets is a statement of the conditions under which sets are formed. In addition to simply listing the members of a set, it was initially assumed that any well-defined condition (or precisely specified property) could be used to determine a set. For example, if $T$ is the property of being a teacup, then the set, $S$, of all teacups might be defined as $S = \{x: T(x)\}$, the set of all individuals, $x$, such that $x$ has the property of being $T$. Even a contradictory property might be used to determine a set. For example, the property of being both $T$ and not $\neg T$ would determine the empty set, the set having no members.

More precisely, naïve set theory assumes the so-called naïve or unrestricted Comprehension Axiom, the axiom that for any formula $\phi(x)$ containing $x$ as a free variable, there will exist the set $\{x: \phi(x)\}$ whose members are exactly those objects that satisfy $\phi(x)$. Thus, if the formula $\phi(x)$ stands for “$x$ is prime”, then $\{x: \phi(x)\}$ will be the set of prime numbers. If $\phi(x)$ stands for “$\sim (x = x)$”, then $\{x: \phi(x)\}$ will be the empty set.

But from the assumption of this axiom, Russell’s contradiction follows. For example, if we let $\phi(x)$ stand for $x \in x$ and let $R = \{x: \sim \phi(x)\}$, then $R$ is the set whose members are exactly those objects that are not members of themselves.

Is $R$ a member of itself? If it is, then it must satisfy the condition of not being a member of itself and so it is not. If it is not, then it must not satisfy the condition of not being a member of itself, and so it must be a member of itself. Since by classical logic one case or the other must hold—either $R$ is a member of itself or it is not—it follows that the theory implies a contradiction.

As Russell tells us, it was after he applied the same kind of reasoning found in Cantor’s diagonal argument to a “supposed class of all imaginable objects” that he was led to the contradiction:

The comprehensive class we are considering, which is to embrace everything, must embrace itself as one of its
members. In other words, if there is such a thing as “everything,” then, “everything” is something, and is a member of
the class “everything.” But normally a class is not a member of itself. Mankind, for example, is not a man. Form
now the assemblage of all classes which are not members of themselves. This is a class: is it a member of itself or
not? If it is, it is one of those classes that are not members of themselves, i.e., it is not a member of itself. If it is not,
it is not one of those classes that are not members of themselves, i.e. it is a member of itself. Thus of the two
hypotheses – that it is, and that it is not, a member of itself – each implies its contradictory. This is a contradiction.  
(1919, 136)

Standard responses to the paradox attempt to limit in some way the conditions under which sets are formed. The goal
is usually both to eliminate $R$ (and similar contradictory sets) and, at the same time, to retain all other sets needed for
mathematics. This is often done by replacing the unrestricted Comprehension Axiom with the more restrictive
Separation Axiom, namely the axiom that given any (consistent) set $S$ and any formula $\phi(x)$ with $x$ free, there will be a set \{ $x \in S: \phi(x)$\} whose members are exactly those members of $S$ that satisfy $\phi(x)$. If we now let $\phi(x)$ stand for the
formula $x \notin x$, it turns out that the corresponding set, \{ $x \in S: x \notin x$\} will not be contradictory since it consists only of
those members found within $S$ that are not members of themselves. Hence the set fails to include itself.

A variety of related paradoxes is discussed in the second chapter of the Introduction to Whitehead and Russell (1910,
2nd edn 60-65), as well as in the entry on paradoxes and contemporary logic in this encyclopedia.

2. History of the Paradox

Russell appears to have discovered his paradox in the late spring of 1901, while working on his Principles of
Mathematics (1903). Exactly when the discovery took place is not clear. Russell initially states that he came across the
paradox “in June 1901” (1944, 13). Later he reports that the discovery took place “in the spring of 1901” (1959, 75). Still
later he reports that he came across the paradox, not in June, but in May of that year (1969, 221). Cesare Burali-Forti,
an assistant to Giuseppe Peano, had discovered a similar antinomy in 1897 when he noticed that since the set of
ordinals is well-ordered, it too must have an ordinal. However, this ordinal must be both an element of the set of all
ordinals and yet greater than every such element.

Unlike Burali-Forti’s paradox, Russell’s paradox does not involve either ordinals or cardinals, relying instead only on
the primitive notions of set and set inclusion. Zermelo noticed a similar contradiction sometime between 1897 and 1902,
possibly anticipating Russell by some years (Ebbinghaus and Peckhaus 2007, 43–48; Tappenden 2013, 336), although
Kanamori concludes that the discovery could easily have been as late as 1902 (Kanamori 2009, 411). As Linsky points
out, Zermelo’s argument, while similar to Russell’s, is best understood as one of a cluster of arguments by Zermelo,
Schröder and Cantor that “indeed anticipated” the mathematical argument Russell developed but that turned out to be
different in small but significant ways from Russell’s argument (Linsky 2013, 11). In any case, the arguments were
thought to be of minor importance until it was realized how detrimental they were to Gottlob Frege’s foundations for
arithmetic.

Russell wrote to Frege with news of his paradox on June 16, 1902. (For the relevant correspondence, see Russell
(1902) and Frege (1902) in van Heijenoort (1967).) The paradox was of significance to Frege’s logical work since, in
effect, it showed that the axioms Frege was using to formalize his logic were inconsistent. Specifically, Frege’s Axiom V
requires that an expression such as $\phi(x)$ be considered both a function of the argument $x$ and a function of the
argument $\phi$. (More precisely, Frege’s Law states that the course-of-values of a concept $f$ is identical to the course-of-
values of a concept $g$ if and only if $f$ and $g$ agree on the value of every argument, i.e., if and only if for every object $x$, $f(x) = g(x)$. See section 2.4.1 of the entry on Gottlob Frege in this encyclopedia for more discussion.) In effect, it was
this ambiguity that allowed Russell to construct $R$ in such a way that it could both be and not be a member of itself.

Russell’s letter arrived just as the second volume of Frege’s Grundgesetze der Arithmetik (The Basic Laws of
Arithmetic, 1893, 1903) was in press. Immediately appreciating the difficulty the paradox posed, Frege added to the
Grundgesetze a hastily composed appendix discussing Russell’s discovery. In the appendix Frege observes that the
consequences of Russell’s paradox are not immediately clear. For example, “Is it always permissible to speak of the
extension of a concept, of a class? And if not, how do we recognize the exceptional cases? Can we always infer from
the extension of one concept’s coinciding with that of a second, that every object which falls under the first concept also
falls under the second? These are the questions,” Frege notes, “raised by Mr Russell’s communication” (1903, 127).
Because of these worries, Frege eventually felt forced to abandon many of his views about logic and mathematics.

Even so, as Russell points out, Frege met the news of the paradox with remarkable fortitude:

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 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 and a telling indication of that of which men are capable if their
dedication is to creative work and knowledge instead of cruder efforts to dominate and be known. (Quoted in van
Heijenoort (1967), 127)

Of course, Russell too was concerned about the consequences of the contradiction. Upon learning that Frege agreed
with him about the significance of the result, he immediately began writing an appendix for his own soon-to-be-released
3. Early Responses to the Paradox

The significance of Russell's paradox can be seen once it is realized that, using classical logic, all sentences follow from a contradiction. For example, assuming both \( P \) and \( \sim P \), any arbitrary proposition, \( Q \), can be proved as follows: from \( P \) we obtain \( P \lor Q \) by the rule of Addition; then from \( P \lor Q \) and \( \sim P \) we obtain \( Q \) by the rule of Disjunctive Syllogism. Because set theory underlies all branches of mathematics, many people began to worry that the inconsistency of set theory would mean that no mathematical proof could be completely trustworthy. Only by eliminating Russell's paradox could mathematics as a whole regain its consistency.

Russell's paradox ultimately stems from the idea that any condition or property may be used to determine a set. For example, the property of being evenly divisible by only itself and the number one distinguishes the set of prime numbers from the set of whole numbers. The property of having mammary glands distinguishes the set of mammals from reptiles, birds and other living organisms. The property of being both square and not square (or any other conjunction of contradictory properties) determines the empty set, and so on.

One early skeptic concerning an unrestricted Comprehension (or Abstraction) axiom was the originator of modern set theory, Georg Cantor. Even prior to Russell's discovery, Cantor had rejected unrestricted Comprehension in favour of what was, in effect, a distinction between sets and classes, recognizing that some properties (such as the property of being an ordinal) produced collections that were simply too large to be sets, and that any assumption to the contrary would lead to inconsistency. (Details can be found in Moore (1982), Hallett (1984) and Menzel (1984).)

Russell's own response to the paradox came with his aptly named theory of types. Believing that self-application lay at the heart of the paradox, Russell's basic idea was that we can avoid commitment to \( R \) (the set of all sets that are not members of themselves) by arranging all sentences (or, more precisely, all propositional functions, functions which give propositions as their values) into a hierarchy. It is then possible to refer to all objects for which a given condition (or predicate) holds only if they are all at the same level or of the same "type."

This solution to Russell's paradox is motivated in large part by adoption of the so-called vicious circle principle. The principle in effect states that no propositional function can be defined prior to specifying the function's scope of application. In other words, before a function can be defined, one must first specify exactly those objects to which the function will apply (the function's domain). For example, before defining the predicate "is a prime number," one first needs to define the collection of objects that might possibly satisfy this predicate, namely the set, \( N \), of natural numbers.

As Whitehead and Russell explain,

An analysis of the paradoxes to be avoided shows that they all result from a kind of vicious circle. The vicious circles in question arise from supposing that a collection of objects may contain members which can only be defined by means of the collection as a whole. Thus, for example, the collection of propositions will be supposed to contain a proposition stating that "all propositions are either true or false." It would seem, however, that such a statement could not be legitimate unless "all propositions" referred to some already definite collection, which it cannot do if new propositions are created by statements about "all propositions." We shall, therefore, have to say that statements about "all propositions" are meaningless. … The principle which enables us to avoid illegitimate totalities may be stated as follows: "Whatever involves all of a collection must not be one of the collection"; or, conversely: "If, provided a certain collection had a total, it would have members only definable in terms of that total, then the said collection has no total." We shall call this the "vicious-circle principle," because it enables us to avoid the vicious circles involved in the assumption of illegitimate totalities. (1910, 2nd edn 37)

If Whitehead and Russell are right, it follows that no function’s scope of application will ever be able to include any object presupposed by the function itself. As a result, propositional functions (along with their corresponding propositions) will end up being arranged in a hierarchy of the kind Russell proposes.

Although Russell first introduced his theory of types in his 1903 Principles of Mathematics, he recognized immediately that more work needed to be done since his initial account seemed to resolve some but not all of the paradoxes. Among the alternatives he considered was a so-called substitutional theory (Galaugher 2013). This in turn led to type theory's more mature expression five years later in Russell's 1908 article, "Mathematical Logic as Based on the Theory of Types," and in the monumental work he co-authored with Alfred North Whitehead, Principia Mathematica (1910, 1912, 1913). Russell's type theory thus appears in two versions: the "simple theory" of 1903 and the "ramified theory" of 1908. Both versions have been criticized for being too ad hoc to eliminate the paradox successfully.

In response to Russell's paradox, David Hilbert also expanded his program of building a consistent, axiomatic foundation for mathematics so that it included an axiomatic foundation for logic and set theory (Peckhaus 2004). Underlying this formalist approach was the idea of allowing the use of only finite, well-defined and constructible objects, together with rules of inference deemed to be absolutely certain.

Finally, Luitzen Brouwer developed intuitionism, whose basic idea was that one cannot assert the existence of a mathematical object unless one can define a procedure for constructing it.

Together, all of these responses helped focus attention on the connections between logic, language and mathematics.
They also helped logicians develop an explicit awareness of the nature of formal systems and of the kinds of metalogical and metamathematical results that have proved to be central to research in the foundations of logic and mathematics over the past one hundred years.

4. Russell’s Paradox in Contemporary Logic

Russell’s paradox is sometimes seen as a negative development – as bringing down Frege’s *Grundgesetze* and as one of the original conceptual sins leading to our expulsion from Cantor’s paradise. W.V. Quine describes the paradox as an “antinomy” that “packs a surprise that can be accommodated by nothing less than a repudiation of our conceptual heritage” (1966, 11). Quine is referring to the Naïve Comprehension principle mentioned earlier. In symbols, the principle states that

\[ \exists A \forall x (x \in A \equiv \phi), \]

where \( A \) is not free in the formula \( \phi \). This says, “There is a set \( A \) such that for any object \( x \), \( x \) is an element of \( A \) if and only if the condition expressed by \( \phi \) holds.” Russell’s paradox arises by taking \( \phi \) to be the formula: \( x \notin x \).

Despite Quine’s comment, it is possible to see Russell’s paradox in a more positive light. For one thing, although the matter remains controversial, later research has revealed that the paradox does not necessarily short circuit Frege’s derivation of arithmetic from logic alone. Frege’s version of NC (his Axiom V) can simply be abandoned. (For details, see the entry on Frege’s Theorem.) For another, Church gives an elegant formulation of the simple theory of types that has proven fruitful even in areas removed from the foundations of mathematics. (For details, see the entry on Type Theory.) Finally, the development of axiomatic (as opposed to naïve) set theories which exhibit various ingenious and philosophically significant ways of dealing with Russell’s paradox paved the way for stunning results in the metamathematics of set theory. These results have included Gödel’s and Cohen’s theorems on the independence of the axiom of choice and Cantor’s continuum hypothesis. So let us see, roughly, how some of these methods – specifically, the so-called “untyped” methods – deal with Russell’s paradox.

Zermelo replaces NC with the following axiom schema of Separation (or *Aussonderungssaxiom*):

\[ \forall A \exists B \forall x (x \in B \equiv (x \in A \land \phi)). \]

Again, to avoid circularity, \( B \) cannot be free in \( \phi \). This demands that in order to gain entry into \( B \), \( x \) must be a member of an existing set \( A \). As one might imagine, this requires a host of additional set-existence axioms, none of which would be required if NC had held up.

How does ZA avoid Russell’s paradox? One might think at first that it doesn’t. After all, if we let \( A \) be \( V \) – the whole universe of sets – and \( \phi \) be \( x \notin x \), a contradiction again appears to arise. But in this case, all the contradiction shows is that \( V \) is not a set. All the contradiction shows is that \( V \) is an empty name (i.e., that it has no reference, that \( V \) does not exist), since the ontology of Zermelo’s system consists solely of sets.

This same point can be made in yet another way, involving a relativized form of Russell’s argument. Let \( B \) be any set. By ZA, the set \( R_B = \{ x \in B : x \notin x \} \) exists, but it cannot be an element of \( B \). For if it is an element of \( B \), then we can ask whether or not it is an element of \( R_B \), and it is if and only if it is not. Thus something, namely \( R_B \), is “missing” from each set \( B \). So again, \( V \) is not a set, since nothing can be missing from \( V \). But notice the following subtlety: unlike the previous argument involving the direct application of *Aussonderungs* to \( V \), the present argument hints at the idea that, while \( V \) is not a set, “\( V \)” is not an empty name. The next strategy for dealing with Russell’s paradox capitalizes on this hint.

John von Neumann’s (1925) untyped method for dealing with paradoxes, and with Russell’s paradox in particular, is simple and ingenious. Von Neumann introduces a distinction between membership and non-membership and, on this basis, draws a distinction between sets and classes. An object is a *member* (simpliciter) if it is a member of some class; and it is a *non-member* if it is not a member of any class. (Actually, von Neumann develops a theory of functions, taken as primitive, rather than classes, wherein corresponding to the member/non-member distinction one has a distinction between an object that can be an argument of some function and one that cannot. In its modern form, due to Bernays and Gödel, it is a single-sorted theory of classes.)

Sets are then defined as members, and non-members are labeled “proper classes.” So for example, the Russell class, \( R \), cannot be a member of any class, and hence it must be a proper class. If \( R \) is assumed to be an element of a class \( A \), then it follows from one of von Neumann’s axioms that \( R \) is not equivalent to \( V \). But \( R \) is equivalent to \( V \), and hence not an element of \( A \). Thus, von Neumann’s method is closely related to the result stated above about the set \( R_B \) for arbitrary \( B \). Von Neumann’s method, while admired by the likes of Gödel and Bernays, has been undervalued in recent years.

Quine (1937) and (1967) similarly provide another untyped method (in letter if not in spirit) of blocking Russell’s paradox, and one that is rife with interesting anomalies. Quine’s basic idea is to introduce a *stratified comprehension axiom*. In effect, the axiom blocks circularity by introducing a hierarchy (or stratification) that is similar to type theory in some ways, and dissimilar in others. (Details can be found in the entry on Quine’s New Foundations.)

In contrast to Zermelo’s, von Neumann’s and Quine’s strategies, which are in a sense purely set theoretic, there have also been attempts to avoid Russell’s paradox by altering the underlying logic. There have been many such attempts
Montague, and Mar (2000). Russell's paradox is an instance of T269 in this list: theorems of pure logic (i.e., of first-order quantification theory with identity)! There is a (partial) list of these in Kalish, It is also worth noting that a number of seemingly purely set-theoretical principles are actually (applied) instances of ironic that although fine-grained Russellian propositions are favored in the philosophy of language, the formal as a foundation for the semantics of natural language. There have also been some recent attempts to obtain the foundations of mathematics, his "other" paradox has yet to lead to anything remotely similar in the foundations of logic of sense and denotation. While we have several set theories to choose from, we do not have anything like a well-developed theory of Russellian propositions, although such propositions are central to the views of Millians and direct-reference theorists. One would think that such a theory would be required for the foundations of semantics, if not for the foundations of mathematics. Thus, while one of Russell's paradoxes has led to the fruitful development of the foundations of mathematics, his "other" paradox has yet to lead to anything remotely similar in the foundations of semantics. To be sure, Church (1974a) and Anderson (1989) have attempted to develop a Russellian intensional logic based on the ramified theory of types, but an argument can be made that the ramified theory is too restrictive to serve as a foundation for the semantics of natural language. There have also been some recent attempts to obtain the beginnings of a Russellian intensional logic based on untyped set theories (Cantini 2004; Deutsch 2014). It is rather ironic that although fine-grained Russellian propositions are favored in the philosophy of language, the formal development of intensional logic is dominated by Montague grammar, with its course-grained theory of propositions. It is also worth noting that a number of seemingly purely set-theoretical principles are actually (applied) instances of theorems of pure logic (i.e., of first-order quantification theory with identity)! There is a (partial) list of these in Kalish, Montague, and Mar (2000). Russell's paradox is an instance of T269 in this list:

\[ (T269) \quad \exists \forall x \exists y (Fxy \equiv \neg Fxx). \]
Reading the dyadic predicate letter “F” as “is a member of”, this says that it is not the case that there is a y such that for any x, x is a member of y if and only if x is not a member of x. Does this mean that Russell’s paradox reduces to T269?

Certainly the proof of T269 distills the essence of Russell’s argument, its pattern of reasoning. But that pattern also underwrites an endless list of seemingly frivolous “pseudo paradoxes” such as the famous paradox of the barber who shaves all and only those who do not shave themselves or, similarly, the paradox of the benevolent but efficient God who helps all and only those who do not help themselves.

How do these “pseudo paradoxes,” as they are sometimes called, differ, if at all, from Russell’s paradox? The pattern of reasoning is the same and the conclusion – that there is no such Barber, no such efficient God, no such set of non-self-membered sets – is the same: such things simply don’t exist. (However, as von Neumann showed, it is not necessary to go quite this far. Von Neumann’s method instructs us not that such things as R do not exist, but just that we cannot say much about them, inasmuch as R and the like cannot fall into the extension of any predicate that qualifies as a class.)

The standard answer to this question is that the difference lies in the subject matter. Quine asks, “why does it [Russell’s paradox] count as an antinomy and the barber paradox not?”; and he answers, “The reason is that there has been in our habits of thought an overwhelming presumption of there being such a class but no presumption of there being such a barber” (1966, 14). Even so, psychological talk of “habits of thought” is not particularly illuminating. More to the point, Russell’s paradox sensibly gives rise to the question of what sets there are; but it is nonsense to wonder, on such grounds as T269, what barbers or Gods there are!

This verdict, however, is not quite fair to fans of the Barber or of T269 generally. They will insist that the question raised by T269 is not what barbers or Gods there are, but rather what non-paradoxical objects there are. This question is virtually the same as that raised by Russell’s paradox itself. Thus, from this perspective, the relation between the Barber and Russell’s paradox is much closer than many (following Quine) have been willing to allow (Salmon 2013).

We note that there is a first-order logical formula that bears the same relation to the principle about the R’s that T269 bears to Russell’s paradox. It is the following:

\[
\forall x \forall y \forall z [Fxy \equiv (Fxz \land \sim Fxx)] \supset \sim Fyz.
\]

(We have taken the liberty of extending the numbering used in Kalish, Montague and Mar (2000) to T273.) But not all set-theoretic paradoxes are similarly related to first-order logical theorems. The Burali-Forti paradox is an example, since the notion of a well-ordering is not elementary; that is, it is not first-order definable.

Russell’s paradox has never been passé, but recently there has been an explosion of interest in it by scholars involved in research in mathematical logic and in philosophical and historical studies of modern logic. A glance at the contents of the 2004 volume One Hundred Years of Russell’s Paradox shows prominent mathematical and philosophical logicians and historians of logic poring over the paradox, proposing new ways back into Cantor’s paradise, or other ways of resolving the issue. Their investigations include radically new ways out of the dilemma posed by the paradox, new studies of the theories of types (simple and ramified, and extensions thereof), new interpretations of Russell’s paradox and constructive theories, of Russell’s paradox of propositions and of his own attempt at an untyped theory (the substitution theory), and so forth.

All of this reminds us that fruitful work can arise from the most unlikely of observations. As Dana Scott has put it, “It is to be understood from the start that Russell’s paradox is not to be regarded as a disaster. It and the related paradoxes show that the naive notion of all-inclusive collections is untenable. That is an interesting result, no doubt about it” (1974, 207).

Bibliography


