LOGIC
Chris Prodoehl

Russell’s Paradox

When people first started to think about sets, they figured there could be sets of anything. Sets are just collections of things. What restrictions could there be on collecting things together, at least in the abstract, and making a set out of them? There is the set of all blue things. There is the set of all movies. There is the set of all movies made before 1953. For any old property you can think of, there is a set of things that have that property. Why not?

As much as this thought might seem like commonsense, Bertrand Russell showed it isn’t right. He did so with what’s come to be known as Russell’s Paradox.

There is more than one way to approach the paradox, though all are closely related to each other. I’ll offer you four. Don’t worry too much about figuring out precisely how each is related to the others.¹

But first two necessary bits of background—two concepts from set theory. First, we say that things are members of sets. A blue dog is a member of the set of blue things, for example. Sets can also be members of themselves. The set of non-blue things is a member of itself, for example. The set of non-blue things isn’t blue: it’s a set, an abstract mathematical object, which has no color, so it is a member of the set of non-blue things. For contrast, notice that the set of blue things is not a member of itself. That set isn’t blue.

Second, sets have subsets in addition to members. Some set S₁ is a subset of another S₂ if and only if every member of S₁ is a member of S₂. Every member of the set of blue dogs is a member of the set of blue things, so the set of blue dogs is a subset of the set of blue things.

The extensions of these two concepts overlap. A subset of some set might also be a member of that set. The set of blue dogs is a subset of the set of blue things, but it isn’t a member of the set of blue things. Again, the set of blue dogs, which is an abstract object, isn’t blue. Compare this with the set of non-blue things. The set of orange dogs is a subset of the set of non-blue things. Every member of the first set is a member of the second. And the set of orange dogs is a member of the set of non-blue things. One more time: the subset isn’t blue, so it is a member of the set of non-blue things.

So we have members of sets, and subsets of sets (where a subset can also be a member!). Now on to the paradox.

¹ The paradox is based on a first-order validity. That validity, in its various instances, implies the existence of various kinds of contradictory entities. This includes a set the existence of which implies a contradiction. My explanation will go in “reverse,” beginning in set theory and ending with the first-order validity.
What’s often called “the Russell set” is defined in terms of the property of self-membership. Or better: non-self-membership. It’s a funny property. For any set S, the Russell set of S is the set of all the members of S that aren’t members of themselves. For example, if the set of blue things is a member of S, then B is a member of the Russell set of S. (Remember, the set of blue things is not a member of itself). In set-builder notation, the Russell set R for any set S is defined as follows:

\[ R = \{ x \mid x \in S \land x \notin x \} \]

Having defined R this way, we can now prove something about it: for any set S, R is a subset of S, but not a member of S. (I.e., all the members of R are members of S, but R is not a member of S.)

Assume for proof by contradiction that R is a member of S. By the law of excluded middle, either R is a member of itself or it isn’t. If it is a member of itself—R∈R—then by definition it fails to be a member of R—R∉R. That’s a contradiction. If R∉R, though, then it satisfies the condition for self-membership, so R∈R. And that’s a contradiction. So, contradiction in either case: whether R∈R or R∉R. We must reject our assumption. The Russell set isn’t a member of S, for any set S.

This result can be used in another proof by contradiction. Consider another funny set: the universal set, V, which contains everything. It’s defined as follows:

\[ V = \{ x \mid x=x \} \]

We can show two things about V: first, that the Russell set is an element of V, and, second, that the Russell set is not an element of V. This is a contradiction.

Either R is or it isn’t an element or V. (By what rule?) It is an element of V because V contains everything. V is the universal set; or, if you like, R=R is true of our set R, so R is an element of V. At the same time, R is not an element of V. Just above we showed that, for any set S (and this includes V), R is not a member of S. So, R is not a member of V. Contradiction.

Since we have a contradiction, we must reject our assumption. Which assumption? That there is a Russell set, i.e., that there is a set defined by the property that defines R—non-self-membership. There is no set of sets-that-aren’t-members-of-themselves.

(You may have noticed that another assumption helps generate the contradiction, namely the assumption that there is a universal set V, which contains everything. Some philosophers do take the lesson of the paradox to be, not that there is no Russell set, but rather that is no universal set. There is no “set of all things.” But we are focusing on what you might call the “standard” interpretation of the paradox here, which takes the Russell set to be the problematic entity.)
There is another way to make the same point—that there can’t be a Russell set—but we have to introduce yet another concept. For any set S, the powerset of S is the set of all S’s subsets. Assuming there is such a thing as the Russell set, R, R is a member of the powerset of S. R for S is the set of all sets in S that aren’t members of themselves. So defined, R is a subset of S, so it is a member of the powerset of S, which, again, is the set of all S’s subsets.

Now we can show that for any set S, its powerset is not a subset of that set. We can do that by showing that there is a member of the powerset that is not a member of S. Remember that R is a member of the powerset of S, and remember we showed in the last section that R is not a member of S. Since not every member of the powerset of S is a member of S, the powerset is not a subset of S.

But we can also show that there is some set such that its powerset is one of its subsets. Here we appeal to the universal set V again. All sets are members of V, including all the subsets in its powerset. But that means there is no member of the powerset of V that is not a member of V. Or, put the other way around, every member of V’s powerset is a member of V. Which is to say that the powerset of V is one of V’s subsets.

This implies a contradiction. We have shown two things. (1) for any set S, the powerset of S is not a subset of S, and (2) the powerset of the universal set V is a subset of V. Therefore, there is a set such that its powerset is a subset of it. Therefore, it is false that, for any set S, the powerset of S is not a subset of S. So we reject the assumption that there is such a thing as the Russell set.

The simplest way to appreciate the paradox is by looking directly at what’s called the Axiom of Unrestricted (or Naïve) Comprehension. Letting the variable “a” range over sets, and the variable “x” range over everything:

\[ \exists a \forall x (x \in a \leftrightarrow P(x)) \] – for any wff P(x), where a doesn’t occur free in P(x).

This says that there is a set such that, for anything at all, it is a member of that set if and only if it has the property P, whatever P is. Then by existential instantiation—letting Q be the relevant set, and replacing P(x) with x \(\notin\) x:

\[ \forall x (x \in Q \leftrightarrow x \not\in x) \]

But the universal quantifier ranges over everything—sets, objects, and the rest; and this includes Q:

\[ (Q \in Q \leftrightarrow Q \not\in Q) \]
This, together with the law of excluded middle, implies a contradiction. So we reject the Axiom of Unrestricted Comprehension. It’s not true that for any property, there is a set of things that have that property. This is tantamount to rejecting the existence of the Russell set, which is that set defined by the property of non-self-membership. We reject the Axiom of Unrestricted Comprehension because it implies the existence of a set whose properties lead to a contradiction. That set is the Russell set, which we rejected above because its properties lead to a contradiction, much like the one we have just derived.

4

In fact, the Axiom of Unrestricted Comprehension can be thought of as an instance of an even more general sentence—one that abstracts away from set theory’s “element” predicate. The contradiction doesn’t depend on the language of set theory, at all—the meaning of “membership,” “element,” even “set.”

Consider the following sentence of first-order logic:

\[ \exists x \forall y (R(x,y) \leftrightarrow \neg R(y,y)) \]

This sentence says that there is something x such that, for every y, x stands in the R-relation to y if and only if y doesn’t stand in the R-relation to itself. This sentence implies the existence of all kinds of contradictory entities—a set whose members are those sets not members of themselves is only one. That’s why Russell tried to explain his insight in various funny ways. There is the barber who shaves only the those who don’t shave themselves (here substituting Shaves(x,y) for R(x,y)). There can be no such barber. Either he shaves himself or he doesn’t. And if he does, then he doesn’t; and if doesn’t, then he does. Or consider the book that references only books that don’t reference themselves (here substituting References(x,y) for R(x,y)). Either the book references itself or it doesn’t. If it does reference itself, then it doesn’t; and if it doesn’t, then it does. There can be no such book!

The following, then, is a first-order validity:

\[-\exists x \forall y (R(x,y) \leftrightarrow \neg R(y,y))\]

And the formal proof of it is not all that complicated. It’s 13.52 in our book.