The Russell paradox (redux)

By **Avi Sion**

This article was first published in *A Fortiori Logic*, by the same author.

Logic is what helps us transmute scattered concrete perceptions into well-ordered abstract concepts. Human knowledge, or opinion, is based on experience, imagination and rational insight. The latter is a kind of 'experience' in the larger sense, a non-phenomenal sort of experience, call it logical 'intuition'. Reason was for this reason called by the ancients, in both West and East, the 'sixth sense' or 'common sense', i.e. the sense-organ which ties together the other five senses, those that bring us in empirical contact with phenomenal experience: colors, shapes, sounds, smells, tastes, touch-sensations, etc., whether they are physically perceived or mentally imagined. The five senses without the sixth yield chaotic nonsense (they are non-sense, one cannot 'make sense' of them); and conversely, the sixth sense is useless without the other five, because it has nothing about which to have rational insights. Imagination reshuffles past experiential data and reasoning, making possible the formation of new ideas and theories which are later tested with reference to further experience and reasoning.

Elements of class logic. Logic initially developed as a science primarily with reference to natural discourse, resulting in what we today refer to as predicate logic. In natural human discourse, we (you and me, and everyone else) routinely think of and discuss things we have perceived, or eventually conceived, by means of categorical propositions involving a subject (say, S) and a predicate (say, P) which are related to each other by means of the copula 'is'. Such propositions have the form "S is P," which may be singular or plural, and in the latter case general (or universal) or particular, and positive or negative, and moreover may involve various modes and categories of modality[1].

A proposition of the form 'S is P' is really a double predication – it tells us that a thing which is S is also P; thus, S and P are really both predicates, though one (the subject S) is given precedence in thought so as to 'predicate' the other (the predicate P) of it[2]. Primarily, S refers to some concrete phenomenon or phenomena (be it/they physical, mental or spiritual), i.e. an individual entity or a set of entities, and P to a property of it or of theirs. For examples, "John is a man" and "All men are human beings" are respectively a singular predication (about one man, John) and a plural one (about all men).

Additionally, still in natural discourse, the subject of our thoughts may be predicates *as such*, i.e. predicates in their capacity as predicates; an example is: "'men' may be the subject or predicate of a proposition." The latter occurs in specifically philosophical (or logical or linguistic) discourse; for example, in the present essay.

Now, logicians through the ages, and especially in modern times, have effectively found natural discourse somewhat inadequate for their needs and gradually developed a more artificial language, that of 'classes'[3]. This type of discourse exactly parallels natural discourse, but is a bit more abstract and descriptive so as to facilitate philosophical (or logical or linguistic) discourse and make it more precise. In this language, instead of saying "this S *is* P," we say "this S *is a member (or instance) of* P" (note well the lengthening of the copula from 'is' to 'is a member (or instance) of'. If 'this S' symbolizes a concrete individual, then 'P' here is called a '*class*'; but if 'this S' symbolizes an abstract class, then 'P' here is called a '*class of classes*'.

A class, then, is an abstraction, a mental constructs in which we figuratively group some concrete things (be they physical, mental or spiritual). Although we can and do temporarily mentally classify things without naming the class for them, we normally name classes (i.e. assign them a distinctive word or phrase) because this facilitates memory and communication. Naming is not the essence of classification, but it is a great facilitator of

---

1 We need not go into the details of these distinctions here, for they are well known. There are also many fine distinctions between different sorts of terms that may appear in propositions as subjects or predicates; but let us keep the matter simple.

2 'Predication' refers to the copula and the predicate together as if they were an action of the speaker (or the statement made) on the subject.

3 The following account of class logic is based on my presentation in *Future Logic*, chapters 43-45. The word 'class' comes from the Latin *classis*, which refers to a "group called to military service" (Merriam-Webster). I do not know whether the Ancients used that word in its logical sense, or some such word, in their discourse, but they certainly thought in class logic mode. Examples of class thinking are Aristotle's distinction between species and genera and Porphyry's tree.

large-scale classification. The name of a class of things does not 'stand for them' in the way of a token, but rather 'points the mind to them' or 'draws our attention to them'; that is to say, it is an instrument of intention.

A class in the primary sense is a class of things in general; a class in the secondary sense is more specifically a class *of classes*. Membership is thus of two kinds: membership of non-classes in a class, or membership of classes in a class of classes. Alternatively, we may speak of first-order classes and second-order classes to distinguish these two types. There are no other orders of classes. When we think about or discuss more concrete things, we are talking in first-order class-logic; when we think about or discuss first-order classes, we are talking in second-order class-logic, and the latter also applies to second-order classes since after all they are classes too. The two orders of classes should not be confused with the hierarchy of classes within each order.

The relation between classes of classes and classes is analogous to the relation between classes and concretes; it is a relation of subsumption. When a lower (i.e. first-order) class is a member of a higher (i.e. second-order) class, it does *not* follow that the members of the lower class are also members of the higher class; in fact, if they are members of the one they are certainly not members of the other. Thus, for example, you and me, although we are members of the class 'men' because we are men, we are not members of the class 'classes of men' because we are not 'men'. Also, the class 'men' is not a man, but is a member of the class 'classes of men'. The members of the class 'classes of men' (or more briefly put, 'men-classes'), which is a class of classes, are, in addition to the broad class 'men', the narrower classes 'gardeners', 'engineers', 'sages', 'neurotics', and so on.[4]

Hierarchization, on the other hand, refers to classes *within a given order* that share instances, not merely by partly overlapping, but in such a way that all the members of one class are members of the other (and in some but not all cases, vice versa). For example, since all men are animals, though not all animals are men, the class 'men' is *a subclass (or species) of* the class 'animals', and the class 'animals' is *an overclass (or genus) of* the class 'men'. If two classes have the same instances, no more and no less, they may be said to be *co-extensive* classes (a class that serves as both species and genus in some context is said to be *sui generis*). If two classes merely share some instances, they may be said to be *intersecting (or overlapping)* classes, but they are not hierarchically arranged (e.g. 'gardeners' and 'engineers'). If two classes of the same order have no instances in common, they may be said to be *mutually exclusive* classes.

It is important to grasp and keep in mind the distinction between hierarchy and order. Since you and I are men, each of us is a member of the class 'men'; this is subsumption by a first-order class of its concrete instances. Since all men are animals, the class 'men' is a subclass of the class 'animals'; this is hierarchy between two classes of the first order. Since 'men' is a class of animals, it is a member of the class 'classes of animals' (or 'animal-classes'); this is subsumption by a second-order class (i.e. a class of classes) of its first-order-class instances (i.e. mere classes). Since all 'classes of men' are 'classes of animals', the class 'men-classes' is a subclass of the class 'animals-classes'; this is hierarchy between two classes of the second order, i.e. between two classes of classes. The relation between classes of the first order and classes of the second order is never one of hierarchy, but always one of subsumption; i.e. the former are always members (instances) of the latter, never subclasses. Hierarchies only occur between classes of the same order.

Thus, in class logic, we have two planes of existence to consider. At the ground level is the relatively objective plane of empirical phenomena (whether these are physical, mental or spiritual in substance); above that, residing in our minds, is the relatively subjective plane of ideas (which are conceived as insubstantial, but do have phenomenal aspects – namely mental or physical images, spoken or written words, and the intentions of such signs), comprising ideas about empirical phenomena and ideas about such ideas. Classes are developed to facilitate our study of empirical phenomena and classes of classes are developed in turn to facilitate our study of classes – for classes (including classes of classes) are of course themselves empirical phenomena of sorts. Classification is a human invention helpful for cognitive ordering of the things observed through our senses or our imaginations or our introspective intuitions. Although classes are products of mind, this does not mean that they are arbitrary – they are formed, organized and controlled by means of our rational faculty, i.e. with the aid of logic.

Clearly, to qualify as a class, a class must have at least one member (in which case the sole member is "one of a kind"). Usually, a class has two or more members, indeed innumerable members. A class is finite if it includes a specified number of instances; if the number of instances it includes is difficult to enumerate, the class is said to be open-ended (meaning infinite or at least indefinite). What brings the instances of a class together in it is their possession of some distinctive property in common; the class is defined by this property (which may of course

---

[4] Note that saying or writing the word men without inverted commas refers to a *predicate*. When we wish to refer to the corresponding *class*, we say the class of men, or the class men; if we are writing, we may write the same with or without inverted commas, or simply 'men' in inverted commas. When dealing with classes of classes, we say the class of classes of men, or the class of men-classes, or the class men-classes, and we may write the same with or without inverted commas, or simply 'classes of men' or 'men-classes' in inverted commas.

be a complicated conjunction of many properties). A class without instances is called a null (or empty) class; this signifies that its defining property is known to be fanciful, so that it is strictly speaking a non-class.

Thus, note well, the term ‘class’ is a bit ambiguous, as it may refer to a first-order class (a class of non-classes, i.e. of things other than classes) or a second-order class (a class of classes, i.e. a mental construct grouping two or more such mental constructs). A class (of the first order) is not, indeed cannot be, a class of classes (i.e. a class of the second order). There is, of course, a class called ‘non-classes’; its instances are principally all concrete things, which are not themselves classes; for example, you and I are non-classes. ‘Non-classes’ is merely a class, not a class of classes, since it does not include any classes. Thus, ‘non-classes’ may be said to be a first-order class, but does not qualify as a second-order class.[5]

The realm of classes of classes is very limited as an object of study in comparison to the realm of mere classes. For what distinctions can we draw between classes? Not many. We can distinguish between classes and classes of classes, between finite and open-ended classes, between positive and negative classes[6], and maybe a few more things, but not much more.

An apparent double paradox. Bertrand Russell (Britain, 1872-1970) proposed a distinction between ‘a class that is a member of itself’ and ‘a class that is not a member of itself’. Although every class is necessarily co-extensive with itself (and in this sense is included in itself), it does not follow that every class is a member of itself (evidently, some are and some are not). Such a distinction can be shown to be legitimate by citing convincing examples. Thus, the class ‘positive classes’ is a member of itself, since it is defined by a positive property; whereas the class ‘negative classes’ is not a member of itself, since it is also positively defined (albeit with general reference to negation). Again, the class ‘finite classes’ is not a member of itself, since it has innumerable members; while the class ‘open-ended classes’ is a member of itself, since it too has innumerable members.

What about the class ‘classes’ – is it a member of itself or not? Since ‘classes’ is a class, it must be a member of ‘classes’ – i.e. of itself. This is said without paying attention to the distinction between classes of the first and second orders. If we ask the question more specifically, the answer has to be nuanced. The class ‘first-order classes’ being a class of classes and not a mere class, cannot be a member of itself, but only a member of ‘second-order classes’; the members of the ‘first-order classes’ are all mere classes. On the other hand, since the class ‘second-order classes’ is a class of classes, it is a member of itself, i.e. a member of ‘second-order classes’. Thus, the class ‘second-order classes’ includes both itself and the class ‘first-order classes’, so that when we say that the wider class ‘classes’ is a member of itself, we mean that it is more specifically a member of the narrower class ‘classes of classes’. As regards, the class ‘non-classes’, since it is a class and not a non-class, it is not a member of itself. Note however that Russell’s paradox does not make a distinction between classes of the first and second orders, but focuses on ‘classes’ indiscriminately.

Russell asked **whether “the class of all classes which are not members of themselves” is or is not a member of itself**. It seemed logically impossible to answer the question, because either way a contradiction ensued. For if the class ‘classes not members of themselves’ *is not* a member of the class ‘classes not members of themselves,’ then it is indeed a member of ‘classes not members of themselves’ (i.e. of itself); and if the class ‘classes not members of themselves’ *is* a member of ‘classes not members of themselves,’ then it is also a member of ‘classes which are members of themselves’ (i.e. of its contradictory). This looked like a mind-blowing double paradox.

The solution of the problem. The pursuit of knowledge is a human enterprise, and therefore one which proceeds by trial and error. Knowledge is inductive much more than deductive; deduction is just one of the tools of induction. There are absolutes in human knowledge, but they are few and far between. When we formulate a theory, it is always essentially a hypothesis, which might later need to be revised or ruled out. So long as it looks useful and sound, and does so more than any competing theory, we adopt it; but if it ever turns out to be belied by some facts or productive of antinomy, we are obliged to either reformulate it or drop it. This is the principle of induction. When we come upon a contradiction, we have to ‘check our premises’ and modify them as necessary. In the case at hand, since our conception of class logic is shown by the Russell paradox to be faulty

[5] Note that, whereas positive terms are easy enough to translate into class logic language, negative terms present a real difficulty. For example, whereas the term men refers only to non-classes, its strict antithesis, the term non-men in its broadest sense, includes both non-classes (i.e. concrete things other than men) and classes (i.e. more abstract things). Again, whereas the term finite classes refers only to classes, its strict antithesis, the term non-finite-classes in its broadest sense, includes both open-ended classes (abstracts) and non-classes (concretes). Thus, we must, for purposes of consistency, admit that some terms do cover both non-classes and classes (including classes of classes). Practically, this means we have to make use of *disjunctives* which reveal the implicit alternatives. This of course complicates class logic considerably.

[6] Positive classes are defined by some positive property and negative classes are defined by a negative one. For examples, ‘men’ is defined with reference to rational animals (positive), whereas ‘bachelors’ is defined with reference to not yet married men (negative).

somehow, we must go back and find out just where we went wrong. So, let us carefully retrace our steps. We defined a class and membership in a class by turning predication into classification, saying effectively:

If something is X, then it is a member of the class ‘X’, and not a member of the class ‘nonX’.

If something is not X, then it is not a member of the class ‘X’, but a member of the class ‘nonX’.

Where did we get this definition? It is not an absolute that was somehow cognitively imposed on us. We invented it – it was a convention by means of which we devised the idea of classes and membership in them. Knowledge can very well proceed without recourse to this idea, and has done so for millennia and continues to do so in many people’s mind. It is an idea with a history, which was added to the arsenal of reasoning techniques by logicians of relatively recent times. These logicians noticed themselves and others reasoning by means of classification, and they realized that this is a useful artifice, distinct from predication and yet based on it somehow. They therefore formally proposed the above definition, and proceeded to study the matter in more detail so as to maximize its utility. The ‘logic of classes’, or ‘class logic’, was born.

However, at some stage, one logician, Bertrand Russell, realized that there was an inherent inconsistency in our conception of classification, which put the whole edifice of class logic in serious doubt. That was the discovery of the paradox bearing his name. That was a great finding, for there is nothing more important to knowledge development, and especially to development of the branch of knowledge called formal logic, than the maintenance of consistency. Every discovery of inconsistency is a stimulation to refine and perfect our knowledge. Russell deserves much credit for this finding, even if he had a lot of difficulty resolving the paradox in a fully convincing manner. Let us here try to do better, by digging deeper into the thought processes involved in classification than he did. What is classification, more precisely?

If we look more closely at our above definition of a class ‘X’ and membership of things in it by virtue of being X, we must ask the question: what does this definition achieve, concretely? Are we merely substituting the phrase ‘is a member of’ for the copula ‘is’, and the class ‘X’ for the predicate X? If this is what we are doing, there is no point in it – for it is obvious that changing *the name* of a relation or a term in no way affects it. Words are incidental to knowledge; what matters is their underlying intent, their meaning. If the words change, but not the meaning, nothing of great significance has changed. No, we are not here merely changing the words used – we are proposing a mental image.

Our idea of classification is that of *mental entities* called classes in which things other than classes (or lesser classes, in the case of classes of classes) are *figuratively collected and contained*. When we say of things that they are members of class ‘X’, we mean that class ‘X’ is a sort of box *into* which these things are, by means of imagination, stored (at a given time, whether temporarily or permanently). That is to say, our ‘definition’ of classification is really a formal convention used to institute this image. What it really means is the following:

If something is X, then it is *in* the class ‘X’, and out of (i.e. *not in*) the class ‘nonX’.

If something is not X, then it is out of (i.e. not in) the class ‘X’, but in the class ‘nonX’.

Clearly, to ‘be’ something and to ‘be in’ (within, inside) something else *are not the same thing*. Our definition conventionally (i.e. by common agreement) decrees that if X is predicated of something, then we may think of that thing as being as if contained by the mental entity called class ‘X’. But this decree is not an absolute; it is not a proposition that being subject to predication of X *naturally and necessarily implies* being a member of class ‘X’. For the whole idea of classification, and therefore this definition of what constitutes a class and membership therein, is a human invention. This invention may well be, and indeed is, very useful – but it remains bound by the laws of nature. If we find that the way we have conceived it, i.e. our definition of it, inevitably leads to contradiction, we must adjust our definition of it in such a way that such contradiction can no longer arise. This is our way of reasoning and acting in all similar situations.

As we shall presently show, since the contradiction is a consequence of the just mentioned defining implication, we must modify that implication. That is to say, we must decree it to have limits. Of course, we cannot just vaguely say that it has limits; we must precisely define these limits so that the practical value of our concept of classification is restored. We can do that by realizing that our definition of classification with reference to something ‘being in’ something else means that *class logic is conceived of as related to geometrical logic*. This is obvious, when we reflect on the fact that we often ‘represent’ classes as geometrical figures (notably, circles) and their members as points within those figures. This practice is not accidental, but of the very essence of our idea of classification. Classification is imagining that we put certain items, identified by their possession of some common and distinctive property, in a labeled container[7].

Let us now examine the concept of self-membership in the light of these reflections. What is the idea of self-membership? It is the presupposition that a class may be a member of itself. But is that notion truly conceivable? If we for a moment put aside the class logic issue, and reformulate the question in terms of geometrical logic, we see that it is absurd. Can a container contain itself? Of course not. There is no known example of a container containing itself in the physical world; and indeed we cannot even visually imagine a container containing itself. So the idea of self-containment has no empirical basis, not even in the mental sphere.

---

[7] This is a pictorial ‘representation’, an analogical image not to be taken literally.

It is only a fanciful conjunction of two words, without experiential basis. For this reason, the idea strikes us as illogical and we can safely posit as a universal and eternal 'axiom' that self-containment is impossible. A nonsensical term like 'the collection of all collections' is of necessity an empty term; we are not forced to accept it, indeed we are logically not allowed to do so; we can only consistently speak of 'the collection of all *other* collections'[8].

A container is of course always co-extensive with itself, i.e. it occupies exactly the space it occupies. But such 'co-extension' is not containment, let alone self-containment, for it does not really (other than verbally) concern two things but only one; there is no 'co-' about it, it is just extended, just once. We refer to containment when a smaller object fits inside a larger object (or in the limit when *another* object of equal size neatly fits inside a certain object). The concept of containment refers to two objects, not one. There has to be two distinct objects; it does not suffice to label the same object in two ways. To imagine 'self-containment' is to imagine that a whole object can somehow fit into itself as a smaller object (or that it can somehow become two, with one of the two inside the other). This is *unconscionable*. A whole thing cannot be a part (whether a full or partial part) of itself; nothing can be both whole and part at once. A single thing cannot be two things (whether of the same or different size) at once; nothing can simultaneously exist as two things.

You cannot decide by convention that something is both whole and part or that one thing is two. You cannot convene something naturally impossible. You can only convene something naturally possible, even though it is unnecessary. Thus, the concept of self-containment is meaningless; it is an inevitably empty concept, because it assumes something impossible to be possible. There is no such thing as self-containment; a container can never contain itself. If this is true, then it is of course equally true that no class includes itself, for (as we have seen) classification is essentially a geometrical idea. Given that a container cannot contain itself, it follows that the answer to the question as to whether a class can be a member of itself is indubitably and definitely: No. Because to say of any class that it is a member of itself is to imply that a container can be a content of itself. Just as no container which is a content of itself exists, so *no class which is a member of itself exists!*

Now, this is a revolutionary idea for class logic. It applies to any and every class, not just to the class 'classes not members of themselves' which gave rise to the Russell paradox. Moreover, note well that we are here denying the possibility of membership of a class in itself, but not the possibility of *non*-membership of a class in itself. When we say that no container contains itself, we imply that it is true of each and every container that it does not contain itself. Similarly, when we say that no class is a member of itself, we imply that it is true of each and every class that it is not a member of itself. What this means is that while we acknowledge the subject of the Russell paradox, namely the class 'classes that are not members of themselves', we reject the notion that such a class might *ever*, even hypothetically for a moment, be a member of itself (and therefore paradoxical) – for, we claim, no class whatever is ever a member of itself.

How can this be, you may well ask? Have we not already shown by example that some classes are members of themselves? Have we not agreed, for example, that the class 'classes' being a class has to be a member of the class 'classes', i.e. of itself? How can we deny something so obvious? Surely, you may well object further, if the class 'classes that are not members of themselves' is not a member of itself, then it is undeniably a member of itself; and if it is a member of itself, then it is undeniably not a member of itself? To answer these legitimate questions, let us go back to our definition of classification, and the things we said about that definition. As I pointed earlier, our definition of classes and membership in them has the form of a conventional implication. It says:

If and only if something is X, then it is a member of the class 'X'.

Now, since this conventional implication leads us inexorably to paradox, we must revise it, i.e. make it more limited in scope, i.e. specify the exact conditions when it 'works' and when it ceases to 'work'. What is essentially wrong with it, as we have seen, is that it suggests that a class can be a member of itself. For example, since the class 'classes' is a class, then it is a member of 'classes'; in this example, the variable X has value class and the variable 'X' has value 'classes'. But, as we have shown, the claim that a class can be a member of itself logically implies something geometrically impossible; namely, that a container can be a content of itself. So, to prevent the Russell paradox from arising, we need to prevent the unwanted consequences of our definition from occurring. Given that our concept of classification is problematic as it stands, what are the conditions we have to specify to delimit it so that the problem is dissolved?

The answer to this question is that when the subject and predicate of the antecedent clause are one and the same, then the consequent clause should cease to be implied. That is to say, if the antecedent clause has the form "if the class 'X' is X" then the consequent clause "then the class 'X' is a member of 'X' (and thus of itself)" *does not follow*. This 'does not follow' is a convention, just as the general 'it follows' was a convention. What we have done here is merely to draw a line, saying that the consequent *generally* follows the antecedent, *except in*

[8] To give a concrete image: a bag of marbles (whether alone or, even worse, with the marbles in it) cannot be put inside itself, even if the bag as a whole, together with all its contents, can be rolled around like a marble and so be called a marble.

the special case where the subject and predicate in the antecedent are 'the same' (in the sense that predicate X is applicable to class 'X' which is itself based on predicate X). This is logically a quite acceptable measure, clearly. If an induced general proposition is found to have exceptions, then it is quite legitimate and indeed obligatory to make it less general, retreating only just enough to allow for these exceptions.

Since the initial definition of classification was a general convention, it is quite permissible, upon discovering that this convention leads us into contradiction, to agree on a slightly narrower convention. Thus, whereas, in the large majority of cases, it remains true that if something is X, then it is a member of the class 'X', and more specifically, if a class (say, 'Y') is X, then it (i.e. 'Y) is a member of the class of classes 'X' – nevertheless, ***exceptionally, in the special case where the class that is X is the class 'X' (i.e. where 'Y' = 'X'), we cannot go on to say of it that it is a member of 'X'***, for this would be to claim it to be a member of itself, which is impossible since this implies that a container can be a content of itself. Note well that we are not denying that, for example, the class 'classes' *is* a class; we are only denying *the implication* this is normally taken to have that the class 'classes' *is a member of* the class 'classes'. We can cheerfully continue saying 'is' (for that is mere predication), but we are not here allowed to turn that 'is' into 'is a member of' (for that would constitute illicit classification).

In this way, the Russell paradox is inhibited from arising. That is to say, with reference to the class 'classes not members of themselves': firstly, it is quite legitimate to suppose that the class 'classes not members of themselves' is not a member of itself, since we know for sure (from geometrical logic) that no class is a member of itself; but it is *not* legitimate to say that this fact (i.e. that it is not a member of itself) implies that it is a member of itself, since such implication has been conventionally excluded. Secondly, it is *not* legitimate to suppose, even for the sake of argument, that the class 'classes not members of themselves' is a member of itself, since we already know (from geometrical logic) that no class is a member of itself, and therefore we cannot establish through such supposition that it is not a member of itself, even though it is anyway true that it is not a member of itself.

As can be seen, our correction of the definition of classification, making it less general than it originally was, by specifying the specific situation in which the implication involved is not to be applied, succeeds in eliminating the Russell paradox. We can say that the class 'classes not members of themselves' is not a member of itself, but we cannot say that it is a member of itself; therefore, both legs of the double paradox are blocked. In the first leg, we have blocked the inference from not-being 'a member of itself' to being one; in the second leg, we have interdicted the supposition of being 'a member of itself' even though inference from it of not-being one would be harmless. Accordingly, the answer to the question posed by Russell – viz. "Is the class of all classes which are not members of themselves a member of itself or not?" – is that this class is not a member of itself, and that this class not-being a member of itself does not, contrary to appearances, make it a member of itself, because no class is a member of itself anyway.

Thus, to be sure, though it is true that the class 'classes' is a class, it does not follow that it is a member of itself; though it is true that the class 'classes of classes' is a class of classes, it does not follow that it is a member of itself; though it is true that the class 'positive classes' is a positive class, it does not follow that it is a member of itself; though it is true that the class 'open-ended classes' is an open-ended class, it does not follow that it is a member of itself; though it is true that the class 'classes that are not members of themselves' is a class that is not a member of itself, it does not follow that it is a member of itself. As for the class 'classes members of themselves', it has no members at all. It should be emphasized that the restriction on classification that we have here introduced is of very limited scope; it hardly affects class logic at all, concerning as it does a few very borderline cases.

The above is, I believe, the *correct and definitive resolution* of the Russell paradox. We acknowledged the existence of a problem, the Russell paradox. We diagnosed the cause of the problem, the assumption that *self-membership* is possible. We showed that self-membership is unconscionable, since it implies that a container can contain itself; this was not arbitrary tinkering, note well, but appealed to reason. We proposed a solution to the problem, one that precisely targets it and surgically removes it. Our remedy consisted in *uncoupling* predication from classification in all cases where self-membership is assumed, and only in such cases. This solution of the problem is plain common sense and not a flight of speculation; it is simple and elegant; it is convincing and uncomplicated; it does not essentially modify the concept of class membership, but only limits its application a little; it introduces a restriction, but one that is clearly circumscribed and quite small; it does not result in collateral damage on areas of class logic, or logic in general, that are not problematic, and therefore does not call for further adaptations of logic doctrine. Note moreover that our solution does not resort to any obscure 'system' of modern symbolic logic, but is entirely developed using ordinary language and widely known and accepted concepts and processes.

A bit of the history. Let us now look briefly at some of the history of the Russell paradox, and see how he and some other modern symbolic logicians dealt with it[9].

Georg Cantor had already in 1895 found an antinomy in his own theory of sets. In 1902, when Gottlob Frege (Germany, 1848-1925) was about to publish the second volume of his *Grundgesetze*, he was advised by Russell of the said paradox. Frege was totally taken in and could not see how to get out of the self-contradictions inherent in "the class of classes that do not belong to themselves." He perceived this as very serious, saying: "What is in question is … whether arithmetic can possibly be given a logical foundation at all." Frege first tried to fix things by suggesting that there might be "concepts with no corresponding classes," or alternatively by adjusting one of his "axioms" in such a way that:

> "Two concepts should be said to have the same extension if, and only if, every object which fell under the first *but was not itself the extension of the first* fell likewise under the second and vice versa"[10].

Clearly, Frege's initial suggestion that there might be "concepts with no corresponding classes" can be viewed as an anticipation of my uncoupling of predication and classification in specific cases. However, Frege did not identify precisely in what cases such uncoupling has to occur. This is evident in his next suggestion, which, though it points tantalizingly to the difficulty in the notion of self-membership, does not reject this notion outright but instead attempts to mitigate it. He speaks of two concepts instead of one, and tries to conventionally exclude the extension as a whole of each from the other, while of course continuing to include the objects falling under the extension; this shows he has not realized that self-inclusion by an extension is not even thinkable.

It should be stressed that Russell's paradox pertains to a certain class (viz. that of all classes not members of themselves) being or not-being a member of *itself* – not of *some other* class. Frege tries to resolve this paradox with reference, not to a single class, but to a pair of equal classes, even though (to my knowledge) he has not demonstrated that co-extensive classes result in a paradox comparable to the Russell paradox. It follows that his attempted solution to the problem is not germane to it. Moreover, Frege seems to have thought that if all items that fall under one class (say, 'Y') fall under another class (say, 'X'), then *the class* 'Y' may be assumed to fall under the class 'X'; and vice versa in the event of co-extension. This is suggested by his attempt to prevent such assumption, so as to avoid (in his estimate) the resulting Russell paradox. But in truth, it does not follow from the given that all Y are X that the class 'Y' is a member of the class 'X' – it only follows that the class 'Y' is a subclass of the class 'X', or an equal class if the relation is reversible. Thus, it appears that Frege confused the relations of class-membership and hierarchization of classes, using a vague term like 'falling' to characterize them both.

We may well ask the question whether an equal class, or a subclass, or even an overclass, might be a member of its hierarchically related class. Offhand, it would seem to be possible. For example, all positive classes are classes and therefore members of the class 'classes', and the class 'positive classes' is a subclass of the class 'classes'; however, although not all classes are positive classes (some are negative classes), nevertheless the class 'classes' is a positive class (being defined by a positive statement), and so is a member of the class 'positive classes'. But although this example suggests that an overclass might be a member of its subclass (and therefore, all the more, an equal class or a subclass might be a member of its hierarchical relative), we might still express a doubt by means of *analogy*, as Frege perhaps intended to do. We could argue inductively, by generalization, that if a class cannot be a member of itself, then maybe it cannot be a member of any coextensive class (as Frege suggests), and perhaps even of a subclass or an overclass. For the issue here is whether the instances referred to by the first class can be thought to occur twice in the second class (as members of it in their own right, and as constituents of a member). So, Frege may have raised a valid issue, which could lead to further restrictions in class logic. However, this need not concern us further in the present context, since (as already explained) it is not directly relevant to resolution of the Russell paradox.

Russell described his paradox in his book *Principles of Mathematics*, published soon after. Although at first inclined to Frege's second approach, he later preferred Frege's first one, proposing that there might be "some propositional functions which did not determine genuine classes." Note here again the failure to pinpoint the precise source and remedy of the problem. Subsequently, Russell thought that "the problem could never be solved completely until all classes were eliminated from logical theory." This, in my view, would be throwing out the baby with the bath water – an overreaction. But then he found out (or rather, he thought he did, or he convinced himself that he did) that the same paradox could be generated without "talk of classes," i.e. with reference to mere predicates – that is, in terms of predicate logic instead of in terms of class logic. As Kneale and Kneale put it (p. 655):

> "Instead of the class which is supposed to contain all classes that are not members of themselves let us consider the property of being a property which does not exemplify itself. If this property exemplifies itself, then it cannot exemplify itself; and if it does not exemplify itself, then it must exemplify itself.

[9] I am here referring principally to the account by William and Martha Kneale in *The Development of Logic* (Oxford, London: Clarendon, 1962), ch. XI.1-2.

[10] Kneale and Kneale, p. 654. Italics theirs.

Clearly, the nature of the trouble is the same here as in the original paradox, and yet there is no talk of classes."

But even if classes are not explicitly mentioned here, it is clear that they are tacitly intended. How would a property 'exemplify' itself? Presumably, property X would be 'a property which exemplifies itself' if property X happens to be one of the things that have property X. That is to say, X exemplifies X if X is a member of the class of things that are X. We cannot talk about properties without resorting to predication; and once we predicate we can (given the initial definition of classification) surely classify. So, this attempt is just verbal chicanery; the same thought is intended, but it is dressed up in other words. It is dishonest. Moreover, the way the paradox is allegedly evoked here does not in fact result in paradox.

We cannot say, even hypothetically, "if this property [i.e. the property of being a property which does not exemplify itself] exemplifies itself" for that is already self-contradictory. To reconstruct a Russell paradox in 'property' terms, we would have to speak of '*the property of all properties* which do not exemplify themselves'; for then we would have a new term to chew on, as we did in class logic. But clearly, this new term is quite contrived and meaningless. Here again, we must mean '*the class of all properties* which do not exemplify themselves' – and in that event, we are back in class logic. Thus, note well, while Russell was right in looking to see whether his paradox was a problem specific to class logic, or one also occurring in predicate logic, and he claimed to have established that it occurred in both fields, in truth (as we have just demonstrated) he did not succeed in doing that. In truth, the paradox *was* specific to class logic; and he would have been better off admitting the fact than trying to ignore it.

In response to certain criticisms by his peers, Russell eventually "agreed that the paradoxes were all due to vicious circles, and laid it down as a principle for the avoidance of such circles that 'whatever involves all of a collection must not be one of the collection'." Thus, Russell may be said to have conceded the principle I have also used, namely that a collection cannot include itself as one of the items collected, although in truth the way he put it suggests he conceived it as a convention designed to block incomprehensible vicious circles rather than a logical absolute (notice that he says 'must not' rather than 'cannot'). He viewed the paradoxes of set theory as "essentially of the same kind as the old paradox of Epimenides (or the Liar)." This suggests that, at this stage, he saw his own paradox as due to self-reference, somehow. It does look at first sight as if there is some sort of self-reference in the proposition 'the class of all classes that are not members of themselves is (or is not) a member of itself', because the clause 'member of itself' is repeated (positively or negatively, in the singular or plural) in subject and predicate[11]. But it cannot be said that self-reference is exactly the problem.

A few years later, in a paper published in 1908, Russell came up with a more elaborate explanation of the Russell paradox based on his 'theory of types'. Russell now argued that "no function can have among its values anything which presupposes the function, for if it had, we could not regard the objects ambiguously denoted by the function as definite until the function was definite, while conversely … the function cannot be definite until its values are definite"[12]. In other words, the question "the class of all classes that are not members of themselves, is it or is it not a member of itself" is inherently flawed, because the subject remains forever out of reach. We cannot take hold of it till we resolve whether or not it is a member of itself, and we cannot do the latter till we do the former; so, the conundrum is unresolvable, i.e. the question is unanswerable. Effectively, the subject is a term cognitively impossible to formulate, due to the double bind the issue of its definition involves for any thinker.

Here, we should note that the purpose of Russell's said explanation was effectively to invalidate the negative class 'classes not members of themselves', since this is the class giving rise to the double paradox he was trying to cure. The positive class 'classes members of themselves' clearly does not result in a double paradox: if we suppose it is *not* a member of itself, self-contradiction does ensue, but we can still say without self-contradiction that it *is* a member of itself. In fact, if Russell's explanation were correct, the positive class ought to be as illicit as the negative one. For if we claim the impossibility of a class referring to something that it *not yet settled*, as Russell did with reference to the negative class, then we must admit this characteristic is also found in the positive class, and we must reject it too. Russell does not seem to have realized that, i.e. that his remedy did not technically differentiate the two classes and so could be applied to both. For this reason, his attempt to solve the Russell paradox with reference to circularity or infinity must be judged as a failure. In my own theory, on the other hand, it is the positive class (that of self-membership) which is invalid (and empty), since it is geometrically unthinkable, while the negative class (that of non-self-membership) remains quite legitimate (and instantiated), as indeed we would expect on the principle that all claims (including that of self-membership) ought to be deniable.

---

[11] Note that if self-reference were the crux of the problem, then the proposition 'the class of all classes that are members of themselves is (or is not) a member of itself' would be equally problematic, even though it apparently does not result in a similar paradox.

[12] Quoted by the Kneales, p. 658.

Anyway, Russell concluded, briefly put, that a function could not be a value of itself; and proposed that function and value be differentiated as two 'types' that could not be mixed together indiscriminately. But this theory is, I would say, too general, and it complicates matters considerably. As we have seen, we cannot refuse to admit that, for instance, 'classes' *is* a class; the most we can do is to deny that this implies that 'classes' *is a member of* itself. This is a denial of self-membership, not of self-predication or of self-reference. As regards the notion of 'types' and later that of 'orders within types', these should not be confused with the more traditional ideas of hierarchies and orders of classes, which we laid out earlier in the present essay. In truth, the resemblance between Russell's concepts and the latter concepts gives Russell's theory a semblance of credibility; but this appearance is quite illusory – these are very different sets of concepts. Russell's notion of 'types' is highly speculative and far from commonsense; while it might appear to solve the Russell paradox, it has ramifications that range far beyond it and incidentally invalidate traditional ideas that do not seem problematic[13]. In short, it is a rough-and-ready, makeshift measure, and not a very convincing one.

Every paradox we come across is, of course, a signal to us that we are going astray somehow. Accordingly, the Russell paradox may be said to have been a signal to Frege, Russell, and other modern logicians, that something was wrong in their outlook. They struggled hard to find the source of the problem, but apparently could not exactly pinpoint its location. All the intricacy and complexity of their symbolic and axiomatic approach to logic could not help them, but rather obscured the solution of the problem for them. This shows that before any attempt at symbolization and axiomatization it is essential for logicians to fully understand the subject at hand in ordinary language terms and by means of commonsense. To my knowledge, the solution of the problem proposed in the present essay is original, i.e. not to be found elsewhere. If that is true, then the theory of class logic developed by modern symbolic logicians, which is still the core of what is being taught in universities today, needs to be thoroughly reviewed and revised.

A bit of self-criticism. As regards the resolution of the Russell paradox that I proposed over two decades ago in my *Future Logic*, the following needs to be said here. While I stand, in the main, by my theory of the logic of classes there (in chapters 43-44), I must now distance myself somewhat from my attempted resolution of the Russell paradox there (in chapter 45).

I did, to my credit, in that past work express great skepticism with regard to the notion of self-membership; but I did not manage to totally rule it out. I did declare: "Intuitively, to me at least, the suggestion that something can be both container and contained is hard to swallow," and I even postulated, in the way of a generalization from a number of cases examined, that "no class of anything, or class of classes of anything, is ever a member of itself," with the possible exception of "things" or "things-classes" (although it might be said of these classes that they are not members of any classes, let alone themselves[14]); but still, I did not reject self-membership on principle, and use that rejection to explain and resolve the Russell paradox, as I do in the present essay.

This is evident, for instance, in my accepting the idea that "'self-member classes' is a member of itself." The reason I did so was the thought that "whether self-membership is possible or not, is not the issue." Superficially, this is of course true – the Russell paradox concerns the 'class of all classes that are not members of themselves', and not 'the class of all classes that are members of themselves'. But in fact, as I have shown today, this is not true; acceptance of self-membership is the true cause of the Russell paradox, and non-self-membership is not in itself problematic.

Anyway, not having duly ruled out self-membership, I resorted to the only solution of the problem that looked promising to me at the time – namely, rejection of 'permutation' from "is (or is not) a member of itself" to "is (or is not) {a member of itself}" (notice the addition of curly brackets). That is to say, I proposed the logical interdiction of changing the *relation* of self-membership or non-self-membership into a predicable *term*. Now I see that this was wrong – it was an action taken *too late* in the process of thought leading up to the Russell paradox. It was a superficial attempt, treating a symptom instead of the disease. I did that, of course, because I thought this was "of all the processes used in developing these arguments, [the] only one of uncertain (unestablished) validity." But in truth, it was not the only possible cause of the effect – there was a process *before* that, one of deeper significance, namely the transition from 'is' to 'is a member of'. I did not at the time notice this earlier process, let alone realize its vulnerability; and for that reason, I did not attack it.

Clearly, I was on the right track, in that I sought for a place along the thought process at which to block development of the Russell paradox. But my error was to pick a place too late along that process. In fact, the right place is earlier on, as advocated in the present essay. The Russell paradox does not arise due to an illicit permutation, but due to the illicit transformation of a predicate into a class in cases where a claim of self-membership would ensue. And while the remedy proposed is even now in a sense 'conventional', the flaw it is designed to fix is quite real – it is that self-membership is in fact impossible and therefore can never be assumed

---

[13] See for a start the Kneales' critique of the 'theory of types' in ch. XI.2.

[14] Note that in this context I come up with the idea that the definition of membership might occasionally fail. But I did not at the time pursue that idea further, because I did not then analyze what such failure would formally imply.

true. My previous proposed solution to the problem only prevented the Russell paradox; it did not prevent self-membership, which is the real cause of the paradox. Thus, the solution I propose in the present essay is more profound and more accurate.

Avi Sion

E-mail: avi-sion@thelogician.net

Links:

- www.TheLogician.net
- The Logician Bookshop
- Avi Sion's Blog
- Avi Sion in Google Books