THE JOURNAL OF SYMBOLIC LOGIC
Volume 48, Number 1, March 1983

PROPER CLASSES

PENELOPE MADDY[1]

§I. Historical introduction. The modern problem of infinity was first raised by Aristotle who held (at least on the popular interpretation[2]) that infinite sets exist potentially (i.e. one more number can always be counted, one more division can always be made in a line segment) but not actually (i.e. the numbers or divisions cannot all exist at one time). In fact, Aristotle not only held that completed infinities never actually exist, but also that they are impossible, that is, that the assumption that they do exist leads to contradictions. To see this, consider Aristotle's view that mathematical entities depend for their existence on the existence of primary substance in which they inhere, coupled with his view that there can be no infinite body. From these it follows that if there were actual completed infinities, infinite bodies would both exist and not exist. The details here are not so important as the idea that if a completed infinite is assumed to exist a contradiction follows.

This negative attitude towards completed infinities flourished for centuries. By the mid-1500's, the German mathematician Stifel (who apparently also invented Pascal's triangle) was moved to condemn irrational numbers simply by their association with the completed infinite:[3]

> ... just as an infinite number is not a number, so an irrational number is not a true number, but lies hidden in a kind of cloud of infinity.

About a hundred years later, Galileo further discredited the completed infinite by pointing out that line segments of different lengths can be brought into one-to-one correspondence by projection, as can the natural numbers and the perfect squares by assigning each number to its square. Here was another paradoxical result of the assumption of completed infinities: sets of clearly different sizes have the same size. All the more reason to ban them from mathematics.

Then came the calculus whose successes were as undeniable as its methods were unacceptable. The main dissenting voice was that of the Irish bishop, George Berkeley, who published (in 1734) a work entitled:

> The Analyst, or a Discourse addressed to an Infidel Mathematician. Wherein it is examined whether the Object, Principles and Inferences of the Modern Analysis are

Received August 10, 1980; revised April 1, 1981.
[1]The author wishes to thank Phillip Bricker for his helpful criticisms.

[2]As often happens, the popular interpretation misses numerous subtleties and controversies. For example, J. Hintikka, *Aristotelian infinity*, ***Philosophical Review,*** vol. 75 (1966), pp. 197-218, and J. Lear, *Aristotelian infinity*, ***Proceedings of the Aristotelian Society,*** vol. 80, pp. 187-210.

[3]Quoted in M. Kline, ***Mathematical thought from ancient to modern times,*** Oxford, 1972, p. 251.

0022-4812/83/4801-0013/$03.70

> more distinctly conceived, or more evidently deduced, than Religious Mysteries and Points of Faith. "First cast out the beam out of thy own eye; and then shalt thou see clearly to cast out the mote out of thy brother's eye."

As the title suggests, Berkeley's aim was to defend faith against the scientific atheists by arguing that the standards of rationality in theology were at least as high as those in the calculus, by this time a cornerstone of science. In the process, he showed that the central notion of an infinitesimal quantity ('*o*' below) was dangerously inconsistent, as in this paraphrase of a Newtonian discussion of the derivative:[4]

> Hitherto I have supposed that '*o*' is something ... From that supposition it is that I get [my conclusions]. I now beg leave to make a new supposition contrary to the first, i.e., I suppose ... that '*o*' is nothing; which second supposition destroys my first, and is inconsistent with it; and therefore with everything that supposeth it. All of which seems a most inconsistent way of arguing, and such as would not be allowed in Divinity.

Thus Berkeley rejected the infinitestimal, the mirror-image of the infinitely large. He was joined by Kant who took the transition from potential to completed infinity as a source of serious error in metaphysics, and probably considered the actual infinite as unnecessary, if not harmful, in mathematics as well. Admidst these attacks on the completed infinite, the calculus remained without a consistent foundation.

Much to everyone's relief, the work of Bolzano, Cauchy and Weierstrass in the late 19th century showed that the infinitely small could be eliminated from the foundations of the calculus, and Berkeley's objections thereby met. But the new theory of limits depended on the theory of real numbers, and the problem of founding the latter was left open. When workable theories of the reals were developed by Dedekind and others, they reintroduced completed infinities.[5]

Thus, the problem of the completed infinite could no longer be avoided. This centuries-old problem was solved by Cantor's work around the turn of the century. Cantor developed a rich theory of transfinite cardinalities in a rough and ready style, and since then the vagaries of his exposition have been eliminated, leaving us with the modern theory of sets. Cantor's theory deals with Galileo's "paradoxical" conclusions by admitting them into the theory in a consistent way: infinite sets which are in one sense of different sizes (one being a proper subset of the other) can still be of the same size in another sense (being in one-to-one correspondence). Aristotle's older contradictory consequence of the existence of completed infinities was eliminated by a philosophical shift from the view that the existence of mathematical entities depends on that of physical instantiations to the view that mathematical entities can exist independently. Cantor himself distinguished two senses in which reality can be ascribed to mathematical concepts: (i) as taking a well-determined place in the understanding, standing in definite relations to other constituents of thought, and (ii) as images or representations of an outer reality of

[4]*The analyst*, paragraph 14, ***The works of George Berkeley, Bishop of Cloyne*** (A. Luce and T. Jessop, eds.), Edinburgh, 1948–1957.

[5]For example, the two sets of rationals which make up a Dedekind cut are both infinite, yet they are considered sufficiently complete to be combined into ordered pairs, then infinite sets of such ordered pairs, and so on.

entities existing independently of human thought. He held that having reality of the first sort implies having reality of the second sort by a difficult metaphysical argument which he did not give, but which depends on "the unity of the All, to which we ourselves belong".[6] The requirements of (i) might seem to leave room for arbitrary choices between equally clear concepts, but Cantor held that in such cases the incorrect concept would betray itself by being inconvenient or unfruitful. Here we have the beginnings of a kind of mathematical realism which depends both on the (almost) self-certifying clarity of some mathematical ideas and on some form of pragmatic or theoretical justification like that of physical science.[7]

Thus ends our rationally reconstructed history. It is supposed to be the story of how the problem of the completed infinite, first raised by Aristotle, aggravated by Galileo and the calculus, was finally solved by Cantor. But does it really show this? What I want to suggest is that just as Cantor had successfully embraced the completed infinities needed for the calculus, the problem of the completed infinite reappeared in his own theory, in a new form, and that Cantor's reaction, though motivated by different considerations, was not unlike Aristotle's. Let me explain.

According to Cantor, corresponding to any set there is a general notion called its cardinal number. These include the usual natural numbers (1, 2, 3, ...) and Cantor's transfinite numbers (for example, $\aleph_0$, the cardinal number of the set of all natural numbers). By showing that the transfinites could be dealt with clearly and rigorously, Cantor introduced completed infinities into mathematical respectability. Continuing in this bold and fearless vein, he next considered the collection of all cardinal numbers. But the assumption that this set exists leads to a contradiction! The problem is (more or less[8]) this: if this set exists, then it has a cardinal number, say κ. Then κ must be bigger than all the cardinal numbers in the set of which it is the cardinal number, that is, in the set of all cardinal numbers. But κ is a cardinal number, so it must be bigger than itself. This is impossible, so there is no set of all cardinal numbers.

Some later results of this sort were quite alarming because it seemed impossible to tell ahead of time which sets could and could not be formed, but Cantor's response was simply to draw a distinction. We have a "consistent multiplicity" or "set" when

> ... the totality of the elements of a multiplicity can be thought of without contradiction as 'being together', so that they can be gathered together into '*one* thing' ... .

---

[6]See P. Jourdain's introduction to Cantor's ***Contributions to the founding of the theory of transfinite numbers,*** Dover, New York, pp. 67–68, and J. Dauben, ***George Cantor,*** Harvard, 1979, pp. 132–133 and the references cited there.

[7]Gödel's mathematical realism is a more developed form of this view. See his *Russell's mathematical logic* and *What is Cantor's continuum problem*? in ***Philosophy of mathematics*** (P. Benacerraf and H. Putnam, eds.), Prentice-Hall, Princeton, N.J., 1964, pp. 211–232, 258–273. I have presented and defended a descendant of Gödel's view *Perception and mathematical intuition*, ***Philosophical Review,*** vol. 89 (1980), pp. 163–196, and in *Sets and numbers*, ***Nous,*** vol. 15(1981), pp.495–511. Both derive from my doctoral dissertation, ***Set theoretic realism***, Princeton University, 1979.

[8]I say "more or less" because this is really Cantor's argument that the set of all ordinal numbers cannot exist. Then he shows that there are as many cardinals as ordinals, and hence that the set of all cardinals cannot exist. See Cantor's *Letter to Dedekind* in ***Mathematical logic from Frege to Gödel*** (J. van Heijenoort, ed.), Harvard University Press, Cambridge, Mass., 1967, pp. 113–117.

On the other hand, when a multiplicity is such that

> . . . the assumption that *all* its elements 'are together' leads to a contradiction, so that it is impossible to conceive the multiplicity as a unity, as 'one finished thing' . . .

then it is an "absolutely infinite" or "inconsistent multiplicity".[9] Notice the similarity between this position and Aristotle's. In both cases, there are objects which are held to exist (the natural numbers, for Aristotle, the finite and transfinite numbers, for Cantor), but it is held that they cannot all exist at the same time (Aristotle) or cannot all exist together (Cantor) on the grounds that the assumption that they do leads to a contradiction.

It should seem surprising that Cantor, the man who had overturned Aristotelianism with respect to certain infinities, would so happily accept it with respect to proper classes (as inconsistent multiplicities are now called[10]). It might be objected that the contradictions Aristotle derived from the existence of small completed infinities depended on questionable metaphysical assumptions, while the contradictions arising from inconsistent multiplicities were purely mathematical in character, but recall that Galileo's paradoxical conclusions were also purely mathematical.

In fact, Cantor's insistence that inconsistent multiplicities be denied the straightforward ontological status he attributed to his transfinite numbers was at least partly based in theology. Pantheism had recently been condemned by papal decree. Theologians worried that Cantor's doctrine of actually existing infinities, combined with doctrines about God's infinity, might lend some support to this heresy. Cantor was a religious man (a Lutheran), and his work on infinity had generally been received with more interest and understanding by theologians than by mathematicians, so he was eager to defend his theory against this charge. His response was that the actual infinite of his transfinite numbers was something much less than the absolute infinite of all transfinite numbers, and that the latter was more proper to God:[11]

> I have never proceeded from any "Genus Supremum" of the actual infinite. Quite the contrary, I have rigorously proven that there is absolutely no "Genus supremum" of the actual infinite. What surpasses all that is finite and transfinite is no "Genus"; it is the single, completely individual unity in which everything is included, which includes the "Absolute", incomprehensible to the human understanding. This is the "Actus Purissimus" which by many is called "God".

This move satisfied the cardinal, and led Cantor to various arguments for his theory of transfinites based on theological considerations.[12]

[9]Cantor, *Letter to Dedekind*, p. 114.

[10]In common usage, a proper class is a collection which for some reason or other cannot be a set. Later I will take up the problem of characterizing the difference between classes generally and sets. Then a proper class will be a class which is not coextensive with any set.

[11]From a letter dated 20 June, 1908 to Grace Chisholm Young, an English mathematician. Cited in J. Dauben, *George Cantor*, ***Isis***, vol. 69 (1978), p. 547.

[12]For a more complete discussion of Cantor's interactions with the church and how they were influenced by Leo XII's encyclical *Aeterni Patris* of 1879, see J. Dauben, *George Cantor*, pp. 140–148.

How should we react to this situation? If we adopt some form of Cantor's mathematical realism, but reject his theological worries about pantheism, there is nothing in this metaphysic to prejudge a search for a consistent mathematical theory of proper classes. In fact, the spirit of this pragmatic realism suggests that the way to decide the question of whether or not proper classes exist is not by prior metaphysical considerations, but by attempting to formulate a consistent theory, and, if successful, by testing that theory for its workability and usefulness. There is room for such testing; modern mathematicians do talk about proper classes. Much of this talk is casual, in the sense that it can be translated away, but in some cases it seems likely that the translated version would never have been reached without the heuristic detour through proper classes. (For example, see Scott's ultrapower construction in the theory of measurable cardinals.[13]) On the other hand, some of the talk is serious, and much in need of foundational clarification. (For example, consider the current interest in reflection arguments.[14]) In general, since the discovery that the axioms of Zermelo and Fraenkel are inadequate to decide the continuum problem, there has been a search for new axioms, and this search has turned up a need for a clearer understanding of the basic notions: set and class. Furthermore, one of the central philosophical difficulties in mathematics concerns the identification of numbers with certain sets. If proper classes were available, they would make more likely, and hopefully less problematic, candidates for such an identification.[15] So, though the need for a theory of proper classes may not be so great now as the need for a theory of limits was in the late 1800's, such a theory would have something to offer both philosophy and the foundations of set theory.

§II. The dilemma posed by various class theories. Cantor's neo-Aristotelian view of proper classes as multiplicities, all of whose members cannot exist together, has already been discussed. A more precisely Aristotelian view would be that the elements cannot all exist at the same time. This suggests a picture of a proper class as a collection whose members are continually coming into existence. This picture is not un-Aristotelian in spirit, but it does offend against our realistic metaphysic which assumes that mathematical entities are generally not the sort of things which come into and go out of existence. Perhaps for this reason, Cantor dropped the temporal reference in his formulation, but he puts nothing in its place. If sets all exist in Cantor's full-blooded sense, we wonder what more is needed for them to exist together. They cannot so exist on pain of contradiction (or so it seems), but what exactly is missing?

---

[13]One begins with V, the class of all sets, then performs a complicated construction using the measurable cardinal. See F. Drake, ***Set theory***, North-Holland, Amsterdam, 1974, Chapter 6, Section 2.

[14]One argues that V, the class of all sets, is "structurally indefinable", and thus that any structural property of V has to be shared by some set. These discussions sometimes involve Ω, the class of all ordinals, and even such things as $\Omega + 1$, and $\Omega + 2$. See W. Reinhardt, *Remarks on reflection principles, large cardinals and elementary embeddings*, ***Axiomatic set theory*** (T. Jech, ed.), American Mathematical Society, Providence, R. I., 1974, pp. 189–205.

[15]See P. Benacerraf, *What numbers could not be*, ***Philosophical Review***, vol. 74 (1965), pp. 47–73. I discuss how proper classes fit into this problem in *Sets and numbers*.

In 1905, a very important step forward in the understanding of the difference between sets and classes was taken by König in an otherwise confused paper which contained a (purported) disproof of the continuum hypothesis.[16] Konig tries to show that the set of real numbers cannot be well-ordered, and hence, that it cannot have cardinality $\aleph_1$. What he actually does is produce a form of definability paradox, but for us the interest of the paper lies elsewhere. After giving his argument, König considers a potential objection, namely that if his argument is correct, then it should apply to any uncountable set, for example to ω_1, which is clearly preposterous. He replies that the word "set" is being used ambiguously:[17]

> When the notion of the continuum is formed, it is the 'arbitrary' sequence $(a_1, a_2, \ldots, a_k, \ldots)$ that is primary, or fundamental. Through the stipulation that $a_1, a_2, \ldots$ are to be replaced by definite positive integers, it becomes a 'definite' sequence, an element of the continuum ... The further stipulation that we consider the *totality* of these 'well-distinguished' objects then leads to the continuum.
>
> The situation is quite different in the case of $[\omega_1]$. Its 'elements' are determined by the 'property' of being order types of well-ordered sets of cardinality $\aleph_0$... this property is only an abstraction, at best a means of distinguishing between objects belonging and objects not belonging to the class; however, it is certainly not a rule according to which *every* element of $[\omega_1]$ can be formed. What is primary, or fundamental, here is the collective notion, which for this very reason, ... I would not call a 'set' but a 'class' ...

König goes on the claim that ω_1 is a class, that is, not a "completed set", but a "set in the process of becoming", and thus that his argument does not apply to it.

In the course of his fertile discussion, König has actually described two forms of the set/class distinction. The second is a version of the neo-Aristotelian set/proper class distinction, but the first, contained in the extended quotation, represents a new approach. This new contrast is the one I want to focus on: the class as the extension of a property versus the set as a combinatorially determined entity. My use of "combinatorial" here follows Bernays:[18]

> These notions are used in a "quasi-combinatorial" sense ... one views a set of integers as the result of infinitely many independent acts deciding for each number whether it should be included or excluded.

This idea of set was developed by Zermelo and became our modern iterative conception:[19] sets are formed in hierarchy of stages. The contrasting idea of classes as extensions of properties, or better, of concepts, formed the groundwork of Frege's theory. On this conception, we imagine the entire universe separated into two heaps depending on whether or not things have the given property or fall under the given

[16]This paper was a sequel to another flawed attempt to disprove the continuum hypothesis which König delivered to the Third International Congress of Mathematicians in 1904. This speech caused Cantor considerable unhappiness and may have contributed to one of his notorious breakdowns. See J. Dauben, *George Cantor*, pp. 543–544. The paper under discussion in the text is *On the foundations of set theory and the continuum problem* in van Heijenoort, op. cit., pp. 145–149.

[17]Ibid., p. 148.

[18]*On Platonism in mathematics* in P. Benacerraf and H. Putnam, op. cit., pp. 275–276.

[19]The first explicit statement of this conception is probably in E. Zermelo, *Uber Grenzzahlen und Mengenbereiche*, ***Fundamenta Mathematicae***, vol. 16 (1930), pp. 29–47. See also G. Kreisel, *Two notes on the foundations of set theory*, ***Dialectica***, vol. 23 (1969), pp. 93–114, and George Boolos, *The iterative conception of set*, ***Journal of Philosophy***, vol. 68 (1971), pp. 215–231.

concept. These contrasting ideas, set as iteratively generated versus class as extension, are often called the mathematical and logical notions of collection, respectively.

Cantor's view can be understood in terms of this contrast. Sets are things which occur in the iterative hierarchy; they are formed in stages. Thus, the elements of a proper class like the class of all sets do all exist, but they do not exist together, in the sense that they do not form a set. The reason for this is now obvious: new sets are formed at each stage, so there cannot be a stage at which the set of all of them is formed. What is missing for this proper class is a stage after which all its elements have been formed.[20]

This Königean version of the set/class distinction is strongly supported. We have seen that it can be traced historically in the works of Zermelo and Frege, and that it helps motivate Cantor's remarks. Furthermore, as Martin has pointed out,[21] it provides an explanation for the controversy over the axiom of choice:

> ... much of the traditional concern about the axiom of choice is probably based on a confusion between sets and definable properties. In many cases it appears unlikely that one can *define* a choice function for a particular collection of sets. But this is entirely unrelated to the question of whether a choice function *exists*. Once this kind of confusion is avoided, the axiom of choice appears as one of the least problematic of the set theoretic axioms.

For these, among other reasons, this version of the set/class distinction is widely adopted in the literature. For example, Gödel remarks that:[22]

> ... this concept of set ... according to which a set is something obtainable from the integers (or some other well-defined objects) by iterated application of the operation 'set of', not something obtained by dividing the totality of existing things into two categories, has never led to any antinomy whatsoever.

Parsons writes:[23]

> ... classes are extensions of predicates ... the conception of an arbitrary subset which appears in set theoretic mathematics is of a *combinatorial* character ...

And finally, Martin:[24]

> ... sets are generated by an iterative construction process. Classes are given all at once, by the properties that determine which objects are members of them ... 'set' is a mathematical concept and 'class' is a logical concept.

As might be anticipated, I intend to embrace this version of the distinction myself; the hard part is seeing how it can be best exploited. What is needed is a system of sets *and* classes which clarifies their differences and their interrelationships. Zermelo and his followers have developed an elegant and workable theory of sets. As far as classes are concerned, the system of Russell and Whitehead in *Principia*

[20]The temporal metaphor is considered inessential. See C. Parsons, *What is the iterative conception of set?* in ***Logic, foundations of mathematics and computability theory*** (Butts and Hintikka, eds.), Reidel, Dordrecht, 1977, pp. 335–367.

[21]D. Martin, *Sets versus classes*, circulated xerox. This was also noted by Bernays, op. cit., p. 276.

[22]*What is Cantor's continuum problem?*, pp. 262–263.

[23]C. Parsons, *Sets and classes*, ***Nous,*** vol. 8 (1974), pp. 7–9.

[24]Martin, op. cit., pp. 9, 11.

Mathematica is the closest descendant of Frege's inconsistent system. It would be interesting, but beside the point, to get into the exact relationship between Frege's ontology and Russell's.[25] The very fact that Russell's is a no-class theory shows that it is not suited to the realistic metaphysic adopted here. Furthermore, the simple theory of types is more appropriate for the mathematical notion, where the sets are constituted by their elements, than for the logical notion, where the entire universe, including both sets and classes, is understood as being divided into two categories. (Zermelo's system is really a simple type theory with cumulative types.) Finally, the ramification brought on by the banning of impredicative definitions (the vicious circle principle) is motivated by a constructivistic picture foreign to our undertaking. These features are enough to show that *PM* is not the system of classes we are looking for, but it might be added that by Cantor's lights, both *PM* and its Quinean descendants[26] should be ruled out by their inconvenience and unfruitfulness for mathematical science. Be that as it may, none of the aforementioned systems treats both sets and classes. Let us turn to an investigation of the conceptual underpinnings of a representative sample of those systems which attempt this task.

Von Neumann reacted to Zermelo's axiomatization of set theory as follows:[27]

> There is, to be sure, a certain justification for the axioms in the fact that they go into evident propositions of naive set theory if in them we take the word 'set', which has no meaning in the axiomatization, in the sense of Cantor. But what is omitted from naive set theory—and to circumvent the antinomies some omission is essential—is absolutely arbitrary.

The system von Neumann proposes to eliminate this arbitrariness can be understood as presupposing two kinds of collections: those which can be elements of other collections, and those which cannot. His motivation for restricting some collections from membership in others is purely the avoidance of paradox. The collections thus restricted are those which are "too big", i.e., those which can be mapped onto the collection of all collections which can be members. Thus, von Neumann allows the "too big" collections disallowed indirectly by Zermelo, but avoids paradox by disallowing their membership in further collections. His axiomatization is elegant: the axiom of "too-bigness" implies Zermelo's separation, Fraenkel's replacement, and a global form of the axiom of choice. Von Neumann had his doubts about this axiom, but in the Cantorian spirit, he justified it pragmatically:[28]

> Axiom IV2, to be sure, requires something more than what was up to now regarded as evident and reasonable for the notion 'not too big'. One might say that it somewhat overshoots the mark. But, in view of the confusion surrounding the notion 'not too big' as it is ordinarily used, on the one hand, and the extraordinary power of this axiom on

[25]There is considerable difficulty involved in simply seeing exactly what Russell's system is, what his ontology includes, and whether or not it does the job it is supposed to do. See Gödel's *Russell's mathematical logic* and W. Quine, *Whitehead and the rise of modern logic* in ***The philosophy of Alfred North Whitehead*** (P. Schilipp, ed.), Northwestern Universty Press, Evanston, 1941, pp. 127–163.

[26]For discussion, see A. Fraenkel, Y. Bar-Hillel, and A. Levy, ***Foundations of set theory,*** North-Holland, Amsterdam, 1973, pp. 161–171.

[27]*An axiomatization of set theory* in van Heijenoort, op. cit., pp. 396.

[28]Ibid., p. 402.

> the other, I believe that I was not too crassly arbitrary in introducing it, especially since it enlarges rather than restricts the domain of set theory, and nevertheless can hardly become a source of antinomies.

How are von Neumann's small and large collections to be understood? The idea is that the small ones are just like Zermelo's sets, and the large ones are entities of exactly the same sort, except that their behavior must be restricted somewhat to avoid paradox. Are the small collections sets and the large collections classes? This could be partly correct, but the large collections cannot be all the classes because some classes are small (e.g. the extension of 'equal to $\varnothing$' is co-extensive with $\{\varnothing\}$, and thus not "too big" in von Neumann's sense). Are the large collections proper classes and the small collections the rest of the classes? No, because the small collections are generated iteratively. Thus, at best, von Neumann's is a distinction between sets and proper classes, just as Cantor's was. As such, it makes some sense on our preferred picture: proper classes are 'too big' in the sense that their members occur arbitrarily high up in the iterative hierarchy, ruling out the possiblilty of a stage after which they have all been formed.

But let us look a little more closely at von Neumann's proper classes. The above considerations do show why these proper classes cannot be elements of sets, but they do not rule out the possibility that proper classes might be elements of other proper classes. In fact, this possibility is strongly indicated by the logical notion. It seems clear that the class of all infinite collections is not only a member of the class of all collections with more than three elements, but also a member of itself. (I shall borrow a notation from Russell and write, for example, $\hat{x}$ (x is infinite) $\in$ $\hat{x}$ (x is infinite).) This indicates that von Neumann's proper classes are not quite our logical ones.

This suspicion that von Neumann's proper classes do not conform to our neo-Königean picture is borne out by the interpretation given to them by later developers of the system. Both Bernays[29] and Gödel[30] view von Neumann's large collections as substitutes for Zermelo's definite properties. Definite properties are those which determine subsets of previously given sets. In modern formulations of Zermelo's axioms, definite properties are replaced by open sentences, but this is considered an unwelcome limitation. As Drake says of the usual first-order version of the separation axiom:[31]

> This axiom can be regarded as an attempt to say that we intend, at each stage of the cumulative structure, to take *every* collection whose members have already been formed as a set at the next level. But we have only formulas of our language with which to describe the collections, and this limits the effect of this axiom.

Zermelo's definite properties were intended to take *every* collection of elements already given; this way of speaking demonstrates that the notion is basically combinatorial. Thus, the use of von Neumann's proper classes as stand-ins for definite properties reinforces my claim that they are more combinatorial than logical.

[29] *A system of axiomatic set theory*, this JOURNAL, vol. 2 (1937), p. 65.

[30] ***The consistency of the continuum hypothesis***, Princeton University Press, Princeton, N. J., 1940, p. 2.

[31] F. Drake, op. cit., p. 9.

This leads to a serious problem in explaining the real nature of von Neumann's proper classes. Our investigation of the picture behind his system leads us to think of the proper classes as combinatorially determined subcollections of the universe of sets. Strengthenings of von Neumann's system like that of Morse and Kelley reinforce this image by striving to increase the number of such subcollections whose existence can be proved. The problem is that when proper classes are combinatorially determined just as sets are, it becomes very difficult to say why this layer of proper classes a top V is not just another stage of sets we forgot to include. It looks like just another rank; saying it is not seems arbitrary. The only difference we can point to is that the proper classes are banned from set membership, but so is the κth rank banned from membership in sets of rank less than κ. Because the classes look so much like just another layer of sets, most set theorists simply think of the proper classes of a weak system like VNBG as metamathematical shorthand, and those of the stronger MK as subsets of a suitably chosen high rank. (For example, take V to be R_κ for κ the first inaccessible cardinal, and take the proper classes to be $\mathscr{P}(R_\kappa) - R_\kappa$.) Let me express the objection I am making here to theories of this sort by saying they draw no significant difference between sets and classes.

Another picture of sets and proper classes is presented by Ackermann's system.[32] It includes an extra constant symbol V for the class of all sets, and its main set comprehension axiom is a reflection principle stating that if φ is a formula not involving V or any proper class parameters, and φ is true only of sets, then there is a set of all sets satisfying φ. The ontological status of Ackermann's proper classes is a bit murky. In defending his restrictions on φ in the reflection schema, he argues that V is forever in the process of becoming, in the process of building, so it cannot be considered "well-defined". The same goes for other proper class parameters. This view is not unfamiliar (it is the neo-Aristotelian view discussed earlier), but it rests uneasily with the fact that Ackermann goes on to allow quantification over classes in φ; it is hard to see how this can be "well-defined" when the individual proper classes are not.[33] Still, Ackermann's system represents an advance over VNBG and MK in one sense: the proper classes are allowed to be elements of other proper classes. (For example, $V \in \hat{x}\,(x = V)$.) Unfortunately though, they are not allowed to be members of themselves, so it seems the only difference is that instead of adding one layer of combinatorially determined proper classes as in MK, Ackermann's system adds several. The natural models of A interpret V as R_κ for some inaccessible κ, and interpret the proper classes as $R_\lambda - R_\kappa$ where $\lambda > \kappa$ and certain definability conditions are satisfied.[34] Reinhardt's system,[35] a descendant of Ackermann's, explicitly assumes regularity for sets and proper classes, reinforcing this picture. Once again, we are faced with the difficult problem of explaining what really distinguishes the layers of proper classes from additional layers of sets.

[32] W. Ackermann, *Zur Axiomatik der Mengenlehre*, ***Mathematische Annalen***, vol. 131 (1956), pp. 336–345. See also A. Levy, *On Ackermann's set theory*, this JOURNAL, vol. 24 (1959), pp. 154–166.

[33] This point is made in A. Fraenkel, Y. Bar-Hillel, and A. Levy, op. cit., p. 150.

[34] See R. Grewe, *Natural models of Ackermann's set theory*, this JOURNAL, vol. 34 (1969), pp. 481–488.

[35] See W. Reinhardt, *Set existence principles of Shoenfield, Ackermann, and Powell*, ***Fundamenta Mathematicae***, vol. 84 (1974), pp. 5–34.

Finally, let me mention Parsons' relativisitic version of the set/proper class distinction.[36] He suggests that the range of my set theoretic quantifier at a given time is some large set as yet "undreamed of" by me, and that the size of this set increases as my theory expands. (This view is actually closer to Lear's;[37] Parsons insists that the range of my quantifier is indeterminate between many such large sets, and that the expansion of my theory just rules out some of these possibilities, or "raises the ante.") Which collections are sets and which are proper classes is relativised to a theorist at a time. A similar view can be expressed in terms of possible worlds.[38]

> ... for a given 'possible world' we should think of the bound variables as ranging over a set, perhaps an R_α; but the sets that exist in that world are elements of the domain, while classes are arbitrary subsets of the domain.

Thus, one world's classes are another world's sets. As Parsons himself notes,[39] this view does not support a real distinction between sets and proper classes. To avoid this problem, Parsons moves toward the idea that a proper class "is not really an object, since even as an intension it is systematically ambiguous."[40] Here he is closer to Cantor, and the official Ackermann, in denying proper classes the full ontological status of sets.

This survey could undoubtedly be extended, but I have enough here to draw my moral. In our search for a realistic theory of sets and classes, we begin with two desiderata:

(1) classes should be real, well-defined entities;

(2) classes should be significantly different from sets.

The central problem is that it is hard to satisfy both of these.[41] Von Neumann, Morse, Kelley and Reinhardt concentrate on (1) and succeed in producing theories with classes as real, well-defined entities, but they run afoul of (2) because their classes look just like additional layers of sets. Lear and Parsons (in the first-mentioned of his moods) again concentrate on (1) and produce real, well-defined proper classes, but these are just sets for another theorist, or for this theorist at another time, or in another possible world, again violating (2). On the other hand, concentrating on (2) leads to Cantor's nonactual, or ineffable proper classes, or the official Ackermann's ill-defined entities, or Parsons's (in his second mood) non-objects. The choice seems to be between a neo-Aristotelianism of ill-defined, potential entities which satisfies (2) but not (1), and some form of a distinction without a difference which satisfies (1) but not (2). Let me now indicate how we might slip between these horns.

---

[36]See C. Parsons, *Sets and classes*, and *What is the iterative conception of set*?

[37]J. Lear, *Sets and semantics*, ***Journal of Philosophy***, vol. 74 (1977), pp. 86–102.

[38]*What is the iterative conception of set*?, p. 350.

[39]Ibid., p. 351.

[40]Ibid., p. 355.

[41]A similar formulation of the central difficulty appears in R. Rucker, *The one/many problem in the foundations of set theory*, ***Logic Colloquium 76***, (R. Grandy and M. Hyland, eds.), North-Holland, Amsterdam, 1977, pp. 567–593.

§III. **A suggestion.** I want a theory of sets and classes based on König's version of the distinction between them. To avoid having this lapse into a distinction without a difference, let us begin by recalling the concrete contrasts this distinction suggests.

First, we have remarked that, pre-theoretically, some extensions seem to be members of others (e.g. $\hat{x}$ (x is infinite) $\in \hat{x}$ (x has more than three elements)), and some extensions even appear to be self-membered (e.g., $\hat{x}$ (x is infinite) $\in \hat{x}$ (x is infinite)). To preserve these intuitions we must allow classes to be elements (as von Neumann, Lear and Parsons would not), and we should not divide them into types (as Russell, Morse, Kelley and Reinhardt would). This last is enough in itself to keep classes distinct from sets, but as a second point of contrast, we should add that any kind of combinatorial determination of classes should be avoided. To put this in deceptive language, imagine all combinatorially determined "subcollections" of V. (This language is deceptive because these do not exist. All combinatorially determined collections are already in V.) We are supposing that classes are closely tied to properties. My point is that we should leave open the question of whether or not there are enough properties to pick out all these "subcollections."

So far, this is an intuitively appealing conception, but unfortunately, it is a version of the one that got Frege into trouble in the first place. As Russell asked, what about $\hat{x}$ ($x \notin x$)? The property '$x \notin x$' seems to divide the world of sets and classes into two categories as the logical notion requires, but the assumption that a class (its extension) corresponds to this property leads to paradox. We seem to have a property without an extension, a property that does not determine a class. Historically, there were two reactions. Zermelo scrapped the logical notion and turned to the mathematical one. Russell tried to retain the logical notion and ended up assuming that a class cannot be of the same type as its elements. This leads to an unacceptable hierarchy of classes.

But there is a third option. To see what it is, consider a similar situation,[42] that is, the problems surrounding such statements as "Everything I've ever said is false". If it turns out that everything I've ever said apart from this statement is false, then the assumption that this statement has a truth value leads to paradox. Here we seem to have a statement without a truth value, where above we had a property without an extension. And again, one solution involves typing. It requires that the idea of truth be typed, so that "Everything I've ever said is false" is a statement of a different type from that of the statements it refers to. (Thus it can be true without being paradoxical.) So, in both cases, we have allowed ourselves to survey a totality (all statements, all classes) which led to a contradiction. And in both cases, this pushed us towards a hierarchy.

Now we are within range of the third option. Kripke[43] has shown how the truth paradoxes can be solved without resort to the Tarskian hierarchy of languages. He does this by allowing truth value gaps as specified by a certain construction.

[42]For discussion of this analogy see C. Parsons, *The liar paradox*, ***Journal of Philosophical Logic***, vol. 3 (1974), pp. 381–412.

[43]S. Kripke, *Outline of a theory of truth*, ***Journal of Philosophy***, vol. 72 (1975), pp. 690–716.

What I propose is that we adapt this solution to the case of logical classes by allowing gaps in the membership relation. To any property, assign an extension and an antiextension, but allow some things to fall in between.

This suggestion is not really new. Some version of this idea might be read into Frege's suggestion that the extensions of two predicates be taken as identical when every object, except the extension itself, falls under the first predicate exactly when it falls under the second.[44] The question of whether or not the extension is self-membered can be thought of as undetermined. Gödel also suggests that the strong medicine of a type hierarchy might be avoided:[45]

> It might turn out that it is possible to assume every concept to be significant everywhere except for certain 'singular points' or 'limiting points', so that the paradoxes appear as something analogous to dividing by zero. Such a system would be most satisfactory in the following respect: our logical intuitions would then remain correct up to certain minor corrections, i.e., they could then be considered to give an essentially correct, only somewhat 'blurred' picture of the real state of affairs.

And finally, Martin makes the suggestion that some sentence (here certain membership sentences) should be allowed to lack truth value:[46]

> Obviously we must make some concession to avoid paradoxes ... The concession I have in mind is that not all sentences will have truth value.

What I propose is to adopt indeterminate membership as a third difference between classes and sets, and to use an imitation of Kripke's construction to show when these indeterminate membership relations occur.

To clarify what will follow, let me review Kripke's construction. He begins with an interpreted first order language L capable of expressing its own syntax, then adds a unary predicate T, which will be only partially defined, to obtain $\mathscr{L}$. A partial definition for T will be an ordered pair (S_1, S_2), where S_1 is the extension of T, and S_2 is its antiextension. S_1 and S_2 must be disjoint, but they need not be exhaustive. $\mathscr{L}(S_1, S_2)$ is an interpreted language with the old interpretations of L plus (S_1, S_2) as the partial interpretation of T. The values assigned to complex formulas are determined by Kleene's strong three-valued logic. (I shall write "$\mathscr{L}(S_1, S_2) \models \varphi$" for "$\mathscr{L}(S_1, S_2)$ thinks φ is true" and "$\mathscr{L}(S_1, S_2) \not\models \varphi$" for "$\mathscr{L}(S_1, S_2)$ thinks φ is false". So, "not $\mathscr{L}(S_1, S_2) \models \varphi$" is not the same as "$\mathscr{L}(S_1, S_2) \not\models \varphi$".) Then

$$
\begin{aligned}
&\mathscr{L}(S_1, S_2) \models \sim\varphi \text{ if } \mathscr{L}(S_1, S_2) \not\models \varphi,\\
&\mathscr{L}(S_1, S_2) \not\models \sim\varphi \text{ if } \mathscr{L}(S_1, S_2) \models \varphi,\\
&\mathscr{L}(S_1, S_2) \models \varphi \vee \psi \text{ if } \mathscr{L}(S_1, S_2) \models \varphi \text{ or } \mathscr{L}(S_1, S_2) \models \psi,\\
&\mathscr{L}(S_1, S_2) \not\models \varphi \vee \psi \text{ if } \mathscr{L}(S_1, S_2) \not\models \varphi \text{ and } \mathscr{L}(S_1, S_2) \not\models \psi,\\
&\mathscr{L}(S_1, S_2) \models \exists x\varphi \text{ if for some } a,\ \mathscr{L}(S_1, S_2) \models \varphi[a],\\
&\mathscr{L}(S_1, S_2) \not\models \exists x\varphi \text{ if for all } a,\ \mathscr{L}(S_1, S_2) \not\models \varphi[a].
\end{aligned}
$$

[44]See G. Frege, ***The basic laws of arithmetic,*** Universiiy of California Press, Berkeley and Los Angeles, California, 1967, translated and edited by M. Furth. p. 139.

[45]*Russell's mathematical logic*, p. 229.

[46]*Sets versus classes*, p. 9.

In all other cases, $\mathscr{L}(S_1, S_2)$ is undecided. The other connectives and the universal quantifier can be defined from these. Thus, $\mathscr{L}(S_1, S_2)$ thinks a conjunction is true if it thinks both conjuncts are true, false if one conjunct is false. And $\mathscr{L}(S_1, S_2)$ thinks $\forall x\varphi$ is true if for all a, $\mathscr{L}(S_1, S_2) \models \varphi[a]$, false if for some a, $\mathscr{L}(S_1, S_2) \not\models \varphi[a]$.

Given any (S_1, S_2), we can form (S_1', S_2') as follows:

$n \in S_1'$ if n is the gödel number of a formula φ such that $\mathscr{L}(S_1, S_2) \models \varphi$;

$n \in S_2'$ if n is the gödel number of a formula φ such that $\mathscr{L}(S_1, S_2) \not\models \varphi$

Then the construction itself goes as follows:

$\mathscr{L}_0 = \mathscr{L}(S_{1,0}, S_{2,0})$ where $S_{1,0} = \varnothing$ and $S_{2,0} = \varnothing$;

$\mathscr{L}_{\alpha+1} = (S_{1,\alpha+1}, S_{2,\alpha+1})$ where $S_{1,\alpha+1} = S_{1,\alpha}'$ and $S_{2,\alpha+1} = S_{2,\alpha}'$;

$\mathscr{L}_\lambda = \mathscr{L}(\bigcup_{\alpha<\lambda} S_{1,\alpha}, \bigcup_{\alpha<\lambda} S_{2,\alpha})$

for limit λ. Simple cardinality considerations show that there must be a countable α for which $\mathscr{L}_\alpha = \mathscr{L}_{\alpha+1}$ and thus $\mathscr{L}_\alpha$ is a language with its own truth predicate.

This construction would never get off the ground if it were not for the following fact:

if $S_1 \subseteq U_1$ and $S_2 \subseteq U_2$, then $S_1' \subseteq U_1'$ and $S_2' \subseteq U_2'$.

In other words, as the construction advances, as new elements are added to the extension and antiextension of T, no truth value previously assigned is changed or becomes undefined. Once a sentence is declared true or false, it never loses that truth value. If the truth value of a complicated statement depends on the truth values of a batch of other statements, its truth value simply will not be determined until the values of the statements in that batch are settled. This is why, for example, "This statement is false "never gets a truth value. It waits forever for the truth values of the statements on which it depends to be determined. This approach makes clear the role of "groundedness" in dealing with the truth theoretic paradoxes.

In the next two sections, I shall imitate Kripke's procedure in a more complex setting to produce a semantics for a language with class terms. While the dependence of Kripke's truth definition on sentences is not an embarrassment, my close linkage between classes and class terms is. I do not mean to suggest that all classes are determined by predicates of any kind, let alone predicates of the language of ZFC, with or without parameters. But the only classes the semantics is able to deal with are the extensions of expressible properties.

§IV. The language $\mathscr{L}$ and its structures. Begin with the language L of ZF, specifically a first order language with one binary predicate letter '$\in$' and without '$=$'. Let

$F_0 =$ The set of all formulas of L with one free variable.

Introduce a new symbol ^ ("hat") and let

$T_0 = \{\hat{x}\varphi / \varphi \in F_0$ and x is the free variable of $\varphi\}$.

So, for example, $\hat{x}\ (x \in x)$ and $\hat{y}\ (\forall z\ (z \notin y))$ are in T_0. (Read: "The class of all x such that $x \in x$," etc.) Now treat the elements of T_0 as constants, and let L_1 be a first order language with one binary predicate '$\in$' and these constants. So for example, $\hat{x}\ (x \in x) \in \hat{x}\ (x \in x)$ and $\forall z\ (z \notin \hat{x}\ (x \in x))$ are formulas of L_1. (N. B. $\forall y(y \in \hat{z}\,(z \in y))$ is not a formula of L_1, because $z \in y$ contains two free variables.) Now let F_1 be the set of all formulas of L_1 with one free variable, and T_1 be the set of all $\hat{x}\varphi$ for $\varphi \in F_1$ and x the free variable of φ. In general, given L_n, F_n and T_n, let

L_{n+1} = a first order language with one binary predicate '$\in$' and the members of T_n as constants.

F_{n+1} = the set of formulas of L_{n+1} with one free variable.

$T_{n+1} = \{\hat{x}\varphi / \varphi \in F_{n+1}$ and x is the free variable of $\varphi\}$.

Finally, let $T = \bigcup T_n$ and let $\mathscr{L}$ be a first order language with one binary predicate '$\in$' and the members of T as constants.

We need to describe an interpretation for the language $\mathscr{L}$. The interpretations of the terms in T will be classes, and we want to allow both classes as members of classes (even self-membership) and cases of indeterminate membership. So, we shall want to interpret a $t \in T$ by an ordered pair (t^+, t^-) where t^+ is thought of as the extension of the class referred to by t, and t^- is the antiextension of that class. (In other words, we represent a class by its extension and antiextension.) Of course, we do not require that all sets and classes be in $t^+ \cup t^-$.

Now obviously, we cannot expect to be able to have the ordered pair (t^+, t^-) itself appear in t^+ or t^- unless we presuppose the kind of theory we are trying to create. But we must allow the possibility that $t \in t$ turns out true. In order to accomplish this, we shall use the term 't' itself as a surrogate for the class it refers to. There is no difficulty with the term 't' appearing in t^+ or t^-.

DEFINITION. An *$\mathscr{L}$-structure* $\mathfrak{A}$ is a nonempty domain of the form $S_{\mathfrak{A}} \cup C_{\mathfrak{A}}$ where each member of $C_{\mathfrak{A}}$ is of the form $(t, t^+_{\mathfrak{A}}, t^-_{\mathfrak{A}})$ for some $t \in T$, and $C_{\mathfrak{A}}$ contains one such triple for each element of T. We also require that

(i) $S_{\mathfrak{A}} \cap C_{\mathfrak{A}} = \varnothing$.

(ii) $t^+_{\mathfrak{A}} \subseteq S_{\mathfrak{A}} \cup T$.

(iii) $t^-_{\mathfrak{A}} \subseteq S_{\mathfrak{A}} \cup T$.

(iv) $t^+_{\mathfrak{A}} \cap t^-_{\mathfrak{A}} = \varnothing$.

(*Notation.* $t^{\mathfrak{A}}$ is the member of $C_{\mathfrak{A}}$ whose first element is t. $(t^{\mathfrak{A}})^+ = t^+_{\mathfrak{A}}$, $(t^{\mathfrak{A}})^- = t^-_{\mathfrak{A}}$, and $T(t^{\mathfrak{A}}) = t$.)

DEFINITION. s is an *$\mathfrak{A}$-sequence* for φ iff s is a function which assigns an element of $\mathfrak{A}$ to each variable of φ and which assigns $t^{\mathfrak{A}}$ to t.

$s(x/a)$ is an $\mathfrak{A}$-sequence exactly like s except that a is assigned to the variable x.

Now we want to define what it is for a structure $\mathfrak{A}$ to think a formula φ is *true* at an $\mathfrak{A}$-sequence s, and what it is for $\mathfrak{A}$ to think a formula *false* at a sequence. (Since $\mathfrak{A}$ will be undecided on some matters, it is not enough just to define truth.) I shall define truth and falsity together by recursion on formulas. The only atomic formulas are of the form $u \in v$ where u and v are terms of $\mathscr{L}$, i.e. where u, v are either variables or members of T. We have the idea that for $t \in T$, the members of the class for which t stands are in $t^+_{\mathfrak{A}}$ and the nonmembers are in $t^-_{\mathfrak{A}}$. This indicates how atomic formulas are to be decided when $s(v) \in C_{\mathfrak{A}}$. But what about the other case, when $s(v) \in S_{\mathfrak{A}}$? Basically, we think of an ordinary set as having its members as members and everything else as nonmembers. This "everything else" includes all classes, since the sets we are concerned with are ordinary sets with sets as members, and no class is equal to any set.

(*Notation.* Write $\mathfrak{A} \not\models \varphi[s]$ for "$\mathfrak{A}$ thinks φ is false at s," not $\mathfrak{A} \models \varphi[s]$ for "It is not the case that $\mathfrak{A}$ thinks φ true at s," and $\mathfrak{A} \models^? \varphi[s]$ for "not $\mathfrak{A} \models \varphi[s]$ and not $\mathfrak{A} \not\models \varphi[s]$.")

DEFINITION. If $\mathfrak{A}$ is an $\mathscr{L}$-structure, φ a formula of $\mathscr{L}$ and s an $\mathfrak{A}$-sequence for φ, then:

(1) If φ is of the form $u \in v$ for terms u, v or $\mathscr{L}$, then
$\mathfrak{A} \models \varphi[s]$ iff:
(i) $s(u), s(v) \in S_{\mathfrak{A}}$ and $s(u) \in s(v)$; or
(ii) $s(u) \in S_{\mathfrak{A}}$ and $s(v) \in C_{\mathfrak{A}}$ and $s(u) \in s(v)^+$; or
(iii) $s(u), s(v) \in C_{\mathfrak{A}}$ and $T(s(u)) = t^*$ and $t^* \in (s(v))^+$.
$\mathfrak{A} \not\models \varphi[s]$ iff:
(i) $s(u), s(v) \in S_{\mathfrak{A}}$ and $s(u) \notin s(v)$; or
(ii) $s(u) \in S_{\mathfrak{A}}$ and $s(v) \in C_{\mathfrak{A}}$ and $s(u) \in s(v)^-$; or
(iii) $s(u), s(v) \in C_{\mathfrak{A}}$ and $T(s(u)) = t^*$ and $t^* \in s(v)^-$; or
(iv) $s(u) \in C_{\mathfrak{A}}$ and $s(v) \in S_{\mathfrak{A}}$.

(2) If φ is of the form $\sim\psi$ then
$\mathfrak{A} \models \varphi[s]$ iff $\mathfrak{A} \not\models \psi[s]$;
$\mathfrak{A} \not\models \varphi[s]$ iff $\mathfrak{A} \models \psi[s]$.

(3) If φ is of the form $\psi \vee \theta$, then
$\mathfrak{A} \models \varphi[s]$ iff $\mathfrak{A} \models \psi[s]$ or $\mathfrak{A} \models \theta[s]$;
$\mathfrak{A} \not\models \varphi[s]$ iff $\mathfrak{A} \not\models \psi[s]$ and $\mathfrak{A} \not\models \theta[s]$.

(4) If φ is of the form $\exists x\psi$, then
$\mathfrak{A} \models \varphi[s]$ iff for some $a \in \mathfrak{A}$, $\mathfrak{A} \models \psi[s(x/a)]$;
$\mathfrak{A} \not\models \varphi[s]$ iff for all $a \in \mathfrak{A}$, $\mathfrak{A} \not\models \psi[s(x/a)]$.

In all other cases, $\mathfrak{A}$ is undecided. (*Notation*. In cases where no confusion will result we write $\mathfrak{A} \models \varphi[{x \atop a}]$ or $\mathfrak{A} \models \varphi[a]$ for $\mathfrak{A} \models \varphi[s(x/a)]$.) Using the standard definitions, we get

PROPOSITION. (5) *If φ is of the form $\psi \wedge \theta$, then*:
$\mathfrak{A} \models \varphi[s]$ *iff* $\mathfrak{A} \models \psi[s]$ *and* $\mathfrak{A} \models \theta[s]$;
$\mathfrak{A} \not\models \varphi[s]$ *iff* $\mathfrak{A} \not\models \psi[s]$ *or* $\mathfrak{A} \not\models \theta[s]$.

(6) *If φ is of the form $\psi \supset \theta$, then*:
$\mathfrak{A} \models \varphi[s]$ *iff* $\mathfrak{A} \not\models \psi[s]$ *or* $\mathfrak{A} \models \theta[s]$;
$\mathfrak{A} \not\models \varphi[s]$ *iff* $\mathfrak{A} \models \psi[s]$ *and* $\mathfrak{A} \not\models \theta[s]$.

(7) *If φ is of the form $\psi \equiv \theta$, then*:
$\mathfrak{A} \models \varphi[s]$ *iff* ($\mathfrak{A} \models \psi[s]$ *and* $\mathfrak{A} \models \theta[s]$) *or* ($\mathfrak{A} \not\models \psi[s]$ *and* $\mathfrak{A} \not\models \theta[s]$);
$\mathfrak{A} \not\models \varphi[s]$ *iff* ($\mathfrak{A} \models \psi[s]$ *and* $\mathfrak{A} \not\models \theta[s]$) *or* ($\mathfrak{A} \not\models \psi[s]$ *and* $\mathfrak{A} \models \theta[s]$).

(8) *If φ is of the form $\forall x\psi$, then*:
$\mathfrak{A} \models \varphi[s]$ *iff for all* $a \in \mathfrak{A}$, $\mathfrak{A} \models \psi[s(x/a)]$;
$\mathfrak{A} \not\models \varphi[s]$ *iff for some* $a \in \mathfrak{A}$, $\mathfrak{A} \not\models \psi[s(x/a)]$.

What happens when new elements are added to $t^+_{\mathfrak{A}}$ or $t^-_{\mathfrak{A}}$ for some term t? We would hope that the resulting structure $\mathfrak{A}'$ would be more decisive than $\mathfrak{A}$ was (i.e. that $\mathfrak{A}'$ would think true (false) some formulas about which $\mathfrak{A}$ was undecided), but not in disagreement with $\mathfrak{A}$ (i.e. $\mathfrak{A}'$ will not think true (false) any formula which $\mathfrak{A}$ thought false (true)). This fact is fundamental.

DEFINITION. If $\mathfrak{A}$, $\mathfrak{A}'$ are $\mathscr{L}$-structures, then $\mathfrak{A} \subsetneq \mathfrak{A}'$ iff $S_{\mathfrak{A}} = S_{\mathfrak{A}'}$ and $t^+_{\mathfrak{A}} \subseteq t^+_{\mathfrak{A}'}$ and $t^-_{\mathfrak{A}} \subseteq t^-_{\mathfrak{A}'}$ for all $t \in T$.

(*Notation*. If s is an $\mathfrak{A}$-sequence, let s' be an $\mathfrak{A}'$-sequence exactly like s except that $s'(t) = t^{\mathfrak{A}'}$ rather than $t^{\mathfrak{A}}$.)

FACT. If $\mathfrak{A} \subsetneq \mathfrak{A}'$, then for all φ and all $\mathfrak{A}$-sequences s for φ:

$\mathfrak{A} \models \varphi[s] \Rightarrow \mathfrak{A}' \models \varphi[s]$;
$\mathfrak{A} \not\models \varphi[s] \Rightarrow \mathfrak{A}' \not\models \varphi[s]$.
The proof is a straightforward induction on the complexity of φ.

§V. **The construction.** What has been done so far can be understood within ordinary model theory, but some features of what follows cannot. For example, I shall need to talk about V, the collection of all sets. Of course, this collection itself is a proper class, so in the metalanguage of a construction meant to clarify the nature of classes, one finds that one must already understand at least one class. (I should also note that I shall consider subclasses of V, or really, or $V \cup T$. It would contradict my own avowed conception of classes if these subcollections were to be understood combinatorially, but in fact they will be tied to terms.) I think this is unavoidable. In other words, to have one's notion of classes clarified, one must already possess a rudimentary notion of class. Notice that the same thing happens when the iterative conception helps us clarify our rudimentary notion of set. I suggest, then, that the following construction be approached as a device to clarify and sharpen our notion of class without reducing it to completely different notions.

To follow Kripke's lead, we want to begin with the standard universe V of sets with the normal membership relation $\in$, and build up the interpretations of the terms T in stages. So, start with $t^+ = t' = \varnothing$ for all $t \in T$, that is, start with the $\mathscr{L}$-structure $\mathfrak{A}_0$ where

$$S_{\mathfrak{A}_0} = V$$

and

$$C_{\mathfrak{A}_0} = \{(t, t_0^+, t_0^-)/t \in T \text{ and } t_0^+ = t_0^- = \varnothing\}.$$

(Notice that t_0^+ and t_0^- are being used as abbreviations for $t_{\mathfrak{A}_0}^+$ and $t_{\mathfrak{A}_0}^-$. It should also be noted that $S_{\mathfrak{A}_0} \cap C_{\mathfrak{A}_0} = \varnothing$ becomes a problem if t is a set. For simplicity, I shall suppose it is not; otherwise, some disjointifying technique would be needed.)

Let us consider a few examples of what $\mathfrak{A}_0$ thinks. It is easy to see that

$$\mathfrak{A}_0 \models x \in y[\varnothing, \{\varnothing\}], \qquad \mathfrak{A}_0 \not\models x \in x[\varnothing].$$

Furthermore, for $t \in T$,

$$\mathfrak{A}_0 \not\models t \in y[\{\varnothing\}], \qquad \mathfrak{A}_0 \models^? x \in t[\varnothing].$$

The next natural questions involve formulas with definite description operators like $x = \varnothing$ where $\varnothing = \urcorner y \forall z(z \notin y)$. Unfortunately $\varnothing$ cannot be picked out in this way because the uniqueness condition is not satisfied.

PROPOSITION. *Not* $\mathfrak{A}_0 \models \forall x \forall x'(\forall z(z \notin x) \wedge \forall z(z \notin x') \supset x = x')$.

PROOF. First recall that $=$ is not a primitive relation of our language $\mathscr{L}$; here it is being used as an abbreviation: $x = x'$ for $\forall w(w \in x \equiv w \in x')$. Now $\mathfrak{A}_0$ thinks the above sentence iff for all $a, b \in \mathfrak{A}_0$, $\mathfrak{A}_0 \not\models \forall z(z \notin x)[a]$ or $\mathfrak{A}_0 \not\models \forall z(z \notin x)[b]$ or $\mathfrak{A}_0 \models x = x'[a, b]$. But $\mathfrak{A}_0 \not\models \forall z(z \notin x)[a]$ iff for some $c \in \mathfrak{A}_0$, $\mathfrak{A}_0 \not\models z \notin x[\begin{smallmatrix} z & x \\ c & a \end{smallmatrix}]$. The trick is that if $a \in C_{\mathfrak{A}_0}$, there is no such c because $(a)_0^+ = \varnothing$. Similarly for the second disjunct if $b \in C_{\mathfrak{A}_0}$. So for this particular choice of a, b, the whole disjunction will fail, unless

$$\mathfrak{A}_0 \models \forall w(w \in x \equiv w \in x')\begin{bmatrix} x & x' \\ a & b \end{bmatrix}.$$

This requires that for all $c \in \mathfrak{A}_0$,

$$\left(\mathfrak{A}_0 \models w \in x \begin{bmatrix} w & x \\ c & a \end{bmatrix} \text{and } \mathfrak{A}_0 \models w \in x' \begin{bmatrix} w & x' \\ c & b \end{bmatrix}\right)$$

or

$$\left(\mathfrak{A}_0 \not\models w \in x \begin{bmatrix} w & x \\ c & a \end{bmatrix} \text{and } \mathfrak{A}_0 \not\models w \in x' \begin{bmatrix} w & x' \\ c & b \end{bmatrix}\right).$$

But neither conjunct of either disjunct is true here because $(a)_0^+ = (a)_0^- = (b)_0^+ = (b)_0^- = \varnothing$. So $\mathfrak{A}_0$ does not think the above sentence is true.

In general, the problem with trying to specify $\varnothing$ uniquely is that there is nothing true of $\varnothing$ which is outright false of any class at this stage, though there are things true of $\varnothing$ which are not true of any class (an example in a moment.) This fact, coupled with our interpretation of $\supset$, makes the mechanism of definite descriptions unusable here. (The cash value of definite descriptions will be recovered when parameters are allowed.)

Despite the failure of the uniqueness condition, $\mathfrak{A}_0$ does distinguish $\varnothing$ from its empty classes in a number of ways. From the abbreviational analysis of $=$ just given, it is easily seen that for all $t \in T$,

$$\text{not } \mathfrak{A}_0 \models x = y\ [\varnothing, t^{\mathfrak{A}_0}].$$

Furthermore, it is worth noting that

$$\mathfrak{A}_0 \models \forall z(z \notin x)\ [\varnothing].$$

But for all $t \in T$,

$$\text{not } \mathfrak{A}_0 \models \forall z(z \notin x)\ [t^{\mathfrak{A}_0}].$$

Now we want to begin adding sets and classes (these latter via their terms) to the t^+'s and t^-'s. Basically if $\mathfrak{A}_0$ thinks φ is true of a set or class, we want to add that set or class to the extension of $\hat{x}\varphi x$. On the other hand, if $\mathfrak{A}_0$ thinks φ is false of a set or class, we want to add it to the antiextension of $\hat{x}\varphi x$. Precisely let $\mathfrak{A}_1$ be an $\mathscr{L}$-structure with

$$S_{\mathfrak{A}_1} = V \quad \text{and} \quad C_{\mathfrak{A}_1} = \{(t, t_1^+, t_1^-) | t \in T\}$$

where for $t = \hat{x}\varphi x$,

$$t_1^+ = \{a \in V \mid \mathfrak{A}_0 \models \varphi[a]\} \cup \{t \in T \mid \mathfrak{A}_0 \models \varphi[t^{\mathfrak{A}_0}]\},$$

$$t_1^- = \{a \in V \mid \mathfrak{A}_0 \not\models \varphi[a]\} \cup \{t \in T \mid \mathfrak{A}_0 \not\models \varphi[t^{\mathfrak{A}_0}]\}.$$

We can verify that $\mathfrak{A}_1$ is an $\mathscr{L}$-structure by noting that $\mathfrak{A}_0$ cannot think any formula both true and false, so $t_1^+ \cap t_1^- = \varnothing$. (This can easily be checked by induction on formulas.) Clearly $\mathfrak{A}_0 \subsetneq \mathfrak{A}_1$, so $\mathfrak{A}_1$ thinks true (false) any formulas $\mathfrak{A}_0$ thought true (false), and it decides some new formulas. For example, by one of our first examples concerning $\mathfrak{A}_0$, we now have

$$\varnothing \in (\hat{x}\ (x \in x))_1^- \quad \text{so} \quad \mathfrak{A}_1 \not\models y \in \hat{x}(x \in x)[\varnothing].$$

And by our penultimate example concerning $\mathfrak{A}_0$, we have

$$\varnothing \in (\hat{x}(\forall z(z \notin x)))_1^+$$

so

$$\mathfrak{A}_1 \models y \in \hat{x}(\forall z(z \notin x))\ [\varnothing].$$

A slightly more involved example:

PROPOSITION. $\mathfrak{A}_1 \models y \in \hat{x}$ (*"x has one element"*) $[\{\varnothing\}]$.

PROOF. Parse "x has one element" as $\exists z(z \in x \wedge \forall w(w \in x \supset w = z))$ where $=$ is understood as above. Then

$$\mathfrak{A}_0 \models \exists z(z \in x \wedge \forall w(w \in x \supset w = z))[\{\varnothing\}]$$

iff for some $a \in \mathfrak{A}_0$,

$$\mathfrak{A}_0 \models z \in x\begin{bmatrix} z & x \\ a & \{\varnothing\}\end{bmatrix} \quad \text{and} \quad \mathfrak{A}_0 \models \forall w(w \in x \supset w = z)[\{\varnothing\}].$$

If we take a to be $\varnothing$, then the first conjunct is easy. The second requires that for all $b \in \mathfrak{A}_0$,

$$\mathfrak{A}_0 \not\models w \in x\begin{bmatrix} w & x \\ b & \{\varnothing\}\end{bmatrix} \quad \text{or} \quad \mathfrak{A}_0 \models w = z\begin{bmatrix} w & z \\ b & \varnothing\end{bmatrix}.$$

The first disjunct holds for all b's except $\varnothing$, and the second holds for $\varnothing$, so $\mathfrak{A}_0$ does think the required formula true at $\{\varnothing\}$. Thus,

$$\{\varnothing\} \in \hat{x}(\text{"}x \text{ has one element"})_1^+,$$

which establishes the proposition.

Let us do one more stage in detail before stating the general form of the transition from $\mathfrak{A}_\alpha$ to $\mathfrak{A}_{\alpha+1}$. Let $\mathfrak{A}_2$ be an $\mathscr{L}$-structure with

$$S_{\mathfrak{A}_2} = V \quad \text{and} \quad C_{\mathfrak{A}_2} = \{(t, t_2^+, t_2^-) | t \in T\},$$

where for $t = \hat{x}\varphi x$,

$$t_2^+ = \{a \in V | \mathfrak{A}_1 \models \varphi[a]\} \cup \{t \in T | \mathfrak{A}_1 \models \varphi[t^{\mathfrak{A}_1}]\},$$
$$t_2^- = \{a \in V | \mathfrak{A}_1 \not\models \varphi[a]\} \cup \{t \in T | \mathfrak{A}_1 \not\models \varphi[t^{\mathfrak{A}_1}]\}.$$

Once again, $\mathfrak{A}_2$ is an $\mathscr{L}$-structure because $\mathfrak{A}_1$ cannot think any formula both true and false. Recall that it was easy to see that $\mathfrak{A}_0 \subsetneq \mathfrak{A}_1$ because $t_0^+ = t_0^- = \varnothing$; this time we need to use the fundamental fact.

PROPOSITION. $\mathfrak{A}_1 \subsetneq \mathfrak{A}_2$.

PROOF. If t is $\hat{x}\varphi x$, then for $a \in V$,

$$a \in t_1^+ \Rightarrow \mathfrak{A}_0 \models \varphi[a] \underset{\mathfrak{A}_0 \subsetneq \mathfrak{A}_1}{\overset{\text{fact}}{\Rightarrow}} \mathfrak{A}_1 \models \varphi[a] \Rightarrow a \in t_2^+.$$

For $u \in T$,

$$u \in t_1^+ \Rightarrow \mathfrak{A}_0 \models \varphi[u^{\mathfrak{A}_0}] \underset{\mathfrak{A}_0 \subsetneq \mathfrak{A}_1}{\overset{\text{fact}}{\Rightarrow}} \mathfrak{A}_1 \models \varphi[u^{\mathfrak{A}_1}] \Rightarrow u \in t_2^+.$$

So $t_1^+ \subseteq t_2^+$. Similarly for $t_1^- \subseteq t_2^-$.

Now let us look at a couple of things that happen at this stage.

PROPOSITION. $\mathfrak{A}_2 \models \hat{x}$(*"x has one element"*) $\in \hat{x}(\exists z(z \in x))$.

PROOF. Recall that $\mathfrak{A}_1 \models y \in \hat{x}$ ("x has one element") $[\{\varnothing\}]$. So

$$\mathfrak{A}_1 \models \exists y(y \in z)\left[\frac{z}{(x(\text{"}x \text{ has one element"}))^{\mathfrak{A}_1}}\right].$$

So

$$\hat{x}(\text{"}x \text{ has one element"}) \in (\hat{z}(\exists y(y \in z)))_2^+.$$

PROPOSITION. $\mathfrak{A}_2 \not\models \hat{z}(\exists x(x \in z)) \in \hat{x}(\forall y(y \in x))$.

PROOF. Clearly

$$\mathfrak{A}_0 \not\models \exists x(x \in z)[\varnothing].$$

So

$$\mathfrak{A}_1 \not\models y \in \hat{z}(\exists x(x \in z))[\varnothing].$$

Thus

$$\mathfrak{A}_1 \not\models \forall y(y \in x)\ [\hat{z}(\exists x(x \in z))^{\mathfrak{A}_1}].$$

So

$$\mathfrak{A}_2 \not\models \hat{z}(\exists x(x \in z)) \in \hat{x}(\forall y(y \in x)).$$

Finally, then, we define $\mathfrak{A}_\alpha$ for ordinals α. $\mathfrak{A}_0$ has been specified. Given $\mathfrak{A}_\alpha$, let $\mathfrak{A}_{\alpha+1}$ be an $\mathscr{L}$-structure with

$$S_{\mathfrak{A}_{\alpha+1}} = V \quad \text{and} \quad C_{\mathfrak{A}_{\alpha+1}} = \{(t, t_{\alpha+1}^+, t_{\alpha+1}^-) | t \in T\}$$

where for $t = \hat{x}\varphi x$,

$$t_{\alpha+1}^+ = \{a \in V | \mathfrak{A}_\alpha \models \varphi[a]\} \cup \{t \in T | \mathfrak{A}_\alpha \models \varphi[t^{\mathfrak{A}_\alpha}]\},$$

$$t_{\alpha+1}^- = \{a \in V | \mathfrak{A}_\alpha \not\models \varphi[a]\} \cup \{t \in T | \mathfrak{A}_\alpha \not\models \varphi[t^{\mathfrak{A}_\alpha}]\},$$

At limit stages λ, let $\mathfrak{A}_\lambda$ be the $\mathscr{L}$-structure with

$$S_{\mathfrak{A}_\lambda} = V \quad \text{and} \quad C_{\mathfrak{A}_\lambda} = \{(t, t_\lambda^+, t_\lambda^-) | t \in T\}$$

where

$$t_\lambda^+ = \bigcup_{\alpha<\lambda} t_\alpha^+ \quad \text{and} \quad t_\lambda^- = \bigcup_{\alpha<\lambda} t_\lambda^-.$$

The following theorem supports the implicit claim that $\mathfrak{A}_\alpha$ is an $\mathscr{L}$-structure for all α, and shows that these $\mathscr{L}$-structures are increasing.

THEOREM. *For all α, $t_\alpha^+ \cap t_\alpha^- = \varnothing$ and for all $\beta \leq \alpha$, $\mathfrak{A}_\beta \subsetneqq \mathfrak{A}_\alpha$.*

PROOF. By induction on α. (1) The result has already been observed for $\alpha = 0, 1$.

(2) Suppose $\alpha = SS\,\gamma$ for some ordinal γ. Then for $t = \hat{x}\varphi x$,

$$t_\alpha^+ = \{a \in V | \mathfrak{A}_{S\gamma} \models \varphi[a]\} \cup \{t \in T | \mathfrak{A}_{s\gamma} \models \varphi[t^{\mathfrak{A}_{s\gamma}}]\},$$

$$t_\alpha^- = \{a \in V | \mathfrak{A}_{S\gamma} \not\models \varphi[a]\} \cup \{t \in T | \mathfrak{A}_{s\gamma} \not\models \varphi[t^{\mathfrak{A}_{s\gamma}}]\}.$$

As before $t_\alpha^+ \cap t_\alpha^- = \varnothing$ because $\mathfrak{A}_{S\gamma}$ cannot think the same formulas both true and false. As for the second part, we already know by the induction hypothesis that for all $\beta \leq S\,\gamma$, $\mathfrak{A}_\beta \subsetneqq \mathfrak{A}_{S\gamma}$. Now $\subsetneqq$ is clearly transitive, so it is enough to show that $\mathfrak{A}_{s\gamma} \subsetneqq \mathfrak{A}_{SS\gamma}$. Here once again we use the fundamental fact. Suppose $t = \hat{x}\varphi x$.

If $a \in V$, then

$$a \in t^{+}_{S\gamma} \Rightarrow \mathfrak{A}_{\gamma} \models \varphi[a] \underset{\text{ind.hyp.}}{\overset{\text{fact}}{\Longrightarrow}} \mathfrak{A}_{S\gamma} \models \varphi[a] \Rightarrow a \in t^{+}_{SS\gamma}.$$

If $u \in T$, then

$$u \in t^{+}_{S\gamma} \Rightarrow \mathfrak{A}_{\gamma} \models \varphi[u^{\mathfrak{A}_{\gamma}}] \underset{\text{ind.hyp.}}{\overset{\text{fact}}{\Longrightarrow}} \mathfrak{A}_{S\gamma} \models \varphi[u^{\mathfrak{A}_{S\gamma}}] \Rightarrow u \in t^{+}_{S\gamma\gamma}.$$

Similarly $t^{-}_{S\gamma} \subseteq t^{-}_{SS\gamma}$. So $\mathfrak{A}_{S\gamma} \subsetneqq \mathfrak{A}_{S\gamma\gamma}$ as required.

(3) Suppose α is a limit ordinal λ. Then $t^{+}_{\lambda} = \bigcup_{\alpha<\lambda} t^{+}_{\alpha}$ and $t^{-}_{\lambda} = \bigcup_{\alpha<\lambda} t^{-}_{\alpha}$. So, if $t^{+}_{\lambda} \cap t^{-}_{\lambda} \neq \varnothing$, then there is an $a \in V \cup T$ such that $a \in \bigcup_{\alpha<\lambda} t^{+}_{\alpha}$ and $a \in \bigcup_{\alpha<\lambda} t^{-}_{\alpha}$. Thus there must be $\beta, \beta' < \lambda$ such that $a \in t^{+}_{\beta}$ and $a \in t^{-}_{\beta}$. Without loss of generality assume $\beta < \beta'$. Then by the induction hypothesis, $t^{+}_{\beta} \subseteq t^{-}_{\beta}$, and thus $a \in t^{+}_{\beta'} \cap t^{-}_{\beta'}$. Contradiction. For the second part, if $\beta < \lambda$ then $\mathfrak{A}_{\beta} \subsetneqq \mathfrak{A}_{\lambda}$ by definition of $\mathfrak{A}_{\lambda}$.

(4) Suppose $\alpha = S\lambda$ for a limit ordinal λ. As before $t^{+}_{S\lambda} \cap t^{-}_{S\lambda} = \varnothing$ because $\mathfrak{A}_{\lambda}$ cannot think the same formula both true and false. For the second part, using the induction hypothesis and the transitivity of $\subsetneqq$, it is enough to show that $\mathfrak{A}_{\lambda} \subsetneqq \mathfrak{A}_{S\lambda}$. Suppose $t = \hat{x}\varphi x$. If $a \in V$, then

$$a \in t^{+}_{\lambda} \Rightarrow a \in \bigcup_{\alpha<\lambda} t^{+}_{\alpha} \Rightarrow \text{there is a } \beta < \lambda \text{ such that } a \in t^{+}_{\beta}$$

$$\underset{\text{hyp.}}{\overset{\text{ind.}}{\Rightarrow}} a \in t^{+}_{\beta+1} \Rightarrow \mathfrak{A}_{\beta} \models \varphi[a] \underset{\text{ind.hyp.}}{\overset{\text{fact}}{\Longrightarrow}} \mathfrak{A}_{\lambda} \models \varphi[a] \Rightarrow a \in t^{+}_{S\lambda}.$$

If $u \in T$, then

$$u \in t^{+}_{\lambda} \Rightarrow u \in \bigcup_{\alpha<\lambda} t^{+}_{\alpha} \Rightarrow \text{there is a } \beta < \lambda \text{ such that } u \in t^{+}_{\beta}$$

$$\underset{\text{hyp.}}{\overset{\text{ind.}}{\Rightarrow}} u \in t^{+}_{\beta+1} \Rightarrow \mathfrak{A}_{\beta} \models \varphi[u^{\mathfrak{A}_{\beta}}] \underset{\text{ind.hyp.}}{\overset{\text{fact}}{\Longrightarrow}} \mathfrak{A}_{\lambda} \models \varphi[u^{\mathfrak{A}_{\lambda}}] \Rightarrow u \in t^{+}_{S\lambda}.$$

Thus $t^{+}_{\lambda} \subseteq t^{+}_{S\lambda}$. Similarly $t^{-}_{\lambda} \subseteq t^{-}_{S\lambda}$, so $\mathfrak{A}_{\lambda} \subsetneqq \mathfrak{A}_{S\lambda}$, as required.

Now we define the final $\mathscr{L}$-structure V^*. Let

$$S_{V^*} = V \quad \text{and} \quad C_{V^*} = \{(t, t^+, t^-) \mid t \in T\}$$

where

$$t^+ = \bigcup_{\alpha} t^{+}_{\alpha} \quad \text{and} \quad t^- = \bigcup_{\alpha} t^{-}_{\alpha}.$$

By imitating the above theorem in the case of a limit ordinal, it can be seen that V^* is an $\mathscr{L}$-structure (i.e. $t^+ \cap t^- = \varnothing$), and that for all α, $\mathfrak{A}_{\alpha} \subsetneqq V^*$. Thus V^* inherits all the beliefs of the $\mathfrak{A}_{\alpha}$'s. For example:

$V^* \models \hat{x}$ ("x has one element") $\in \hat{x}(\exists z(z \in x))$.

$V^* \not\models \hat{z}(\exists x(x \in z)) \in \hat{x}(\forall y(y \in x))$.

One example of self-membership is surely in order:

Proposition. $V^* \models \hat{z}(\exists x(x \in z)) \in \hat{z}(\exists x(x \in z))$.

Proof. We noted earlier that

$$\mathfrak{A}_2 \models \hat{x} \text{ ("}x\text{ has one element") } \in \hat{x}(\exists z(z \in x)).$$

Thus

$$\mathfrak{A}_2 \models \exists z(z \in \hat{x}(\exists z(z \in x))).$$

So

$$\hat{x}(\exists z(z \in x)) \in \hat{x}(\exists z(z \in x))^{+}_{3}.$$

So

$$\mathfrak{A}_3 \models \hat{x}(\exists z(z \in x)) \in \hat{x}(\exists z(z \in x)).$$

And finally since $\mathfrak{A}_3 \subsetneqq V^*$, the proposition is established.

Here is one question that is never decided:

PROPOSITION. $V^* \models \hat{x}(x \notin x) \in \hat{x}(x \notin x)$.

PROOF. Suppose $V^* \models \hat{x}(x \notin x) \in \hat{x}(x \notin x)$. Then

$$\hat{x}(x \notin x) \in \hat{x}(x \notin x)^+ = \bigcup_\alpha \hat{x}(x \notin x)^+_\alpha.$$

So, there is a least α such that $\hat{x}(x \notin x) \in \hat{x}(x \notin x)^+_\alpha$. α cannot be limit, so there is a γ such that $\alpha = S\gamma$ and

$$\mathfrak{A}_\gamma \models x \notin x[\hat{x}(x \notin x)^{\mathfrak{A}_\gamma}].$$

So

$$\mathfrak{A}_\gamma \models \hat{x}(x \notin x) \notin \hat{x}(x \notin x).$$

But $\mathfrak{A}_\gamma \subsetneqq \mathfrak{A}_\alpha$, so

$$\mathfrak{A}_\alpha \models \hat{x}(x \notin x) \notin \hat{x}(x \notin x) \quad \text{and} \quad \mathfrak{A}_\alpha \models \hat{x}(x \notin x) \in \hat{x}(x \notin x),$$

which is impossible. Similarly, it cannot be the case that $V^* \not\models \hat{x}(x \notin x) \in \hat{x}(x \notin x)$.

Finally a remark about equality. Recall that the language $\mathscr{L}$ does not contain '='. We can now expand to a language with '=' for V^*. Let the constants of $\mathscr{L}_=$ be the terms of T, and the binary predicates be '$\in$' and '='. Interpret the elements of T and '$\in$' as before, and let $V^* \models x = y\,[a, b]$ iff (i) $a, b \in V$ and $a = b$, or (ii) $a, b \in C_{V^*}$ and $(a)^+ = (b)^+$ and $(a)^- = (b)^-$. Then it is easy to see that

PROPOSITION. *For all a, $b \in V^*$ and all formulas φ of $\mathscr{L}_=$,*

(i) $V^* \models x = x[a]$,

(ii) *if* $V^* \models x = y\,[a, b]$, *then* $V^* \models \varphi[a] \Rightarrow V^* \models \varphi[b]$ *and* $V^* \not\models \varphi[a] \Rightarrow V^* \not\models \varphi[b]$.

(Notice that since the terms of T are constants, '=' will not appear within the scope of an $\hat{x}$.)

§VI. A nontrivial example. I want to establish the following:

THEOREM. $V^* \models \hat{x}(\text{"}x \textit{ is infinite}\text{"}) \in \hat{x}(\text{"}x \textit{ is infinite}\text{"})$.

To simplify the argument, let us pause to consider this notion:

DEFINITION. φ is *setbounded* in $\mathfrak{A}$ at s iff

(i) φ contains no class terms,

(ii) range $s \subseteq V$,

(iii) all φ's quantifiers are bounded (i.e. occur in the form $\forall x(x \in y \supset \cdots)$ or $\exists x(x \in y \wedge \cdots)$ where x and y are different variables).

PROPOSITION. *If φ is setbounded in $\mathfrak{A}_\alpha$ at s, then*

$\mathfrak{A}_\alpha \models \varphi[s]$ *iff φ is true in V at s,*

$\mathfrak{A}_\alpha \not\models \varphi[s]$ *iff φ is false in V at s.*

PROOF. By induction on the complexity of φ.

Case 1. If φ is atomic then it is of the form $u \in v$ where u, v are variables (because φ contains no class terms). Then

$\mathfrak{A}_\alpha \models \varphi[s]$ iff $\mathfrak{A}_\alpha \models u \in v[s]$ iff $s(u) \in s(v)$ (because $s(u)$, $s(v) \in V$) iff φ is true in V at s.

Similarly,

$\mathfrak{A}_\alpha \not\models \varphi[s]$ iff $s(u) \notin s(v)$ iff φ is false in V at s.

Cases 2 *and* 3. φ is of the form $\sim\psi$ or $\psi \vee \theta$. These are perfectly straightforward.

Case 4. φ is of the form $\exists x\psi$. Then ψ is of the form $x \in y \wedge \theta$ where x is different from y and θ is setbounded, so

$$\begin{aligned}
\mathfrak{A}_\alpha \models \varphi[s] \text{ iff } & \mathfrak{A}_\alpha \models \exists x(x \in y \wedge \theta)[s] \\
& \text{iff there is an } a \in \mathfrak{A}_\alpha \text{ such that } \mathfrak{A}_\alpha \models x \in y \wedge \theta\,[s(x/a)] \\
& \text{iff there is an } a \in \mathfrak{A}_\alpha \text{ such that} \\
& \quad \mathfrak{A}_\alpha \models x \in y[s(x/a)] \text{ and } \mathfrak{A}_\alpha \models \theta[s(x/a)] \\
& \text{iff (because } s(y) \in V\text{) there is an } a \in \mathfrak{A}_\alpha \text{ such that} \\
& \quad a \in s(y) \in V \text{ and } \mathfrak{A}_\alpha \models \theta[s(x/a)] \\
& \underset{\text{hyp.}}{\overset{\text{ind.}}{\text{iff}}} \text{ (because } a \in V\text{) there is an } a \in \mathfrak{A}_\alpha \text{ such that} \\
& \quad a \in s(y) \in V \text{ and } \theta \text{ is true in } V \text{ at } s(x/a) \\
& \text{iff there is an } a \in \mathfrak{A}_\alpha \text{ such that} \\
& \quad x \in y \wedge \theta \text{ is true in } V \text{ at } s(x/a) \\
& \text{iff } \varphi \text{ is true in } V \text{ at } s.
\end{aligned}$$

$$\begin{aligned}
\mathfrak{A}_\alpha \not\models \varphi[s] \text{ iff } & \mathfrak{A}_\alpha \not\models \exists x(x \in y \wedge \theta)[s] \\
& \text{iff for all } a \in \mathfrak{A}_\alpha,\ \mathfrak{A}_\alpha \not\models x \in y \wedge \theta[s(x/a)] \\
& \text{iff for all } a \in \mathfrak{A}_\alpha,\ \mathfrak{A}_\alpha \not\models x \in y[s(x/a)] \text{ or } \mathfrak{A}_\alpha \not\models \theta[s(x/a)] \\
& \text{iff (because if } a \in C_{\mathfrak{A}_\alpha},\ \mathfrak{A} \not\models x \in y[s(x/a)]) \\
& \quad \text{for all } a \in V,\ \mathfrak{A}_\alpha \not\models x \in y[s(x/a)] \text{ or } \mathfrak{A}_\alpha \not\models \theta[s(x/a)]. \\
& \underset{\text{hyp.}}{\overset{\text{ind.}}{\text{iff}}} \text{ for all } a \in V,\ a \notin s(y) \in V \text{ or } \theta \text{ is false in } V \text{ at } s(x/a) \\
& \text{iff for all } a \in V,\ x \in y \text{ is false in } V \text{ at } s(x/a) \\
& \quad \text{or } \theta \text{ is false in } V \text{ at } s(x/a) \\
& \text{iff } \exists x(x \in y \wedge \theta) \text{ is false in } V \text{ at } s \\
& \text{iff } \varphi \text{ is false in } V \text{ at } S.
\end{aligned}$$

COROLLARY. *If φ is setbounded in $\mathfrak{A}_\alpha$ at S, then not $\mathfrak{A}_\alpha \models^? \varphi[s]$.*

To return to the theorem, let "x is infinite" be

$$\exists y\Big(y \subseteq x \wedge \exists z\Big(z \subsetneq y \wedge \exists f\colon z \xrightarrow[\text{onto}]{\text{1-1}} y\Big)\Big).$$

(This actually says that x contains an infinite subcollection. This eliminates difficulties about the existence of functions with proper classes as range.)

I shall divide the proof into two main lemmas.

LEMMA 1. *If A is the set of all infinite subsets of N, and $a \in A$, then $\mathfrak{A}_0 \models$ "x is infinite"* $[a]$.

If this lemma were established then we would have

$$\mathfrak{A}_1 \models y \in \hat{x}(\text{“}x \text{ is infinite”})[a]$$

so

$$\mathfrak{A}_1 \models \forall w(w \in y \supset w \in \hat{x}(\text{“}x \text{ is infinite”}))[A]$$

or

$$\mathfrak{A}_1 \models y \subseteq x(\text{“}x \text{ is infinite”})[A].$$

LEMMA 2. $\mathfrak{A}_1 \models \exists z(z \subsetneq y \wedge \exists f: z \xrightarrow[\text{onto}]{\text{1-1}} y)[A]$.

If this lemma were established, then we would have

$$\mathfrak{A}_1 \models \text{“}x \text{ is infinite”} \ [\hat{x}(\text{“}x \text{ is infinite”})^{\mathfrak{A}_1}]$$

and thus

$$\hat{x}(\text{“}x \text{ is infinite”}) \in \hat{x}(\text{“}x \text{ is infinite”})_2^+$$

so

$$\mathfrak{A}_2 \models \hat{x}(\text{“}x \text{ is infinite”}) \in \hat{x}(\text{“}x \text{ is infinite”}).$$

Finally, $\mathfrak{A}_2 \subsetneq V^*$, so the theorem would be established.

PROOF OF LEMMA 1. If $a \in A$, then a *is* infinite, so there is a $b \subsetneq a$ and a function

$$F: b \xrightarrow[\text{onto}]{\text{1-1}} a.$$

We shall break the lemma into a series of claims:

(i) $\mathfrak{A}_0 \models y \subseteq x[a, a]$: $y \subseteq x$ is $\forall w(w \in y \supset w \in x)$. This formula is setbounded in $\mathfrak{A}_0$ at a and true in V at a.

(ii) $\mathfrak{A}_0 \models z \subsetneq y \ [b, a]$: $z \subsetneq y$ is $\forall w(w \in z \supset w \in y) \wedge \exists w(w \in y \wedge w \notin z)$. This is setbounded in $\mathfrak{A}_0$ at $[b, a]$ and true in V at $[b, a]$.

(iii) $\mathfrak{A}_0 \models$ "x is a function" $[F]$: "x is a function" is "x is a relation" and "x is functional". "x is a relation" is

$$\forall z(z \in x \supset \exists u \exists v(u \in z \wedge v \in z \wedge \exists s \exists t(s \in u \wedge s \in v \wedge t \in v$$
$$\wedge \forall w(w \in u \supset w = s) \wedge \forall w(w \in v \supset (w = s \vee w = t)))))$$

where $w = s$ is $\forall y(y \in w \supset y \in s) \wedge \forall y(y \in s \supset y \in w)$. ('$=$' will be so understood throughout this argument.) Similarly for $w = t$. This formula is setbounded in $\mathfrak{A}_0$ at F and true in V at F. "x is functional" is

$$\forall y \forall z(y \in x \wedge z \in x \supset (\exists u(u \in y \wedge u \in z \wedge \exists v$$
$$(v \in u \wedge \forall w(w \in u \supset w = v))) \supset y = z)).$$

Again this formula is setbounded in $\mathfrak{A}_0$ at F and true in V at F.

(iv) $\mathfrak{A}_0 \models$ "x is 1-1" $[F]$: "x is 1-1" is

$$\forall y \forall z(y \in x \wedge z \in x \supset (\exists u(\exists v \exists v'(v \in y \wedge v' \in y \wedge u \in v \wedge u \notin v') \\ \wedge \exists v \exists v'(v \in z \wedge v' \in z \wedge u \in v \wedge u \notin v')) \supset y = z)).$$

Again, this formula is setbounded in $\mathfrak{A}_0$ at F and true in V at F.

(v) $\mathfrak{A}_0 \models$ "dom $x = y$" $[F, b]$: "dom $x = y$" is "dom $x \subseteq y$" and "$y \subseteq$ dom x", or

$$\forall z(z \in x \supset \exists w(w \in z \wedge \exists u(u \in w \wedge \forall v(v \in w \supset v = u)) \\ \wedge \forall s(s \in w \supset s \in y))) \\ \wedge \forall z(z \in y \supset \exists u(u \in x \wedge \exists v(v \in u \wedge z \in v \wedge \forall w(w \in v \supset w = z)))).$$

This formula is setbounded in $\mathfrak{A}_0$ at $[F, b]$ and true in V at $[F, b]$.

(vi) $\mathfrak{A}_0 \models$ "range $x = y$" $[F, a]$: "range $x = y$" is "range $x \subseteq y$" and "$y \subseteq$ range x", or

$$\forall z(z \in x \supset \exists u \exists v(u \in z \wedge v \in z \wedge \exists s \exists t \\ (s \in u \wedge s \in v \wedge t \in v \wedge t \in y))) \\ \wedge \forall z(z \in y \supset \exists u(u \in x \wedge \exists s \exists t(s \in u \wedge t \in u \wedge \\ \exists n \exists m(n \in s \wedge n \in t \wedge m \in t \wedge m = z)))).$$

This formula is setbounded in $\mathfrak{A}_0$ at $[F, a]$ and true in V at $[F, a]$. Applying the proposition to (i)–(vi), we have

$$\mathfrak{A}_0 \models y \subseteq x \wedge z \subsetneq y \wedge f\colon z \xrightarrow[\text{onto}]{\text{1-1}} y \begin{bmatrix} y & x & z & f \\ a & a & b & F \end{bmatrix}.$$

So

$$\mathfrak{A}_0 \models \exists y\Big(y \subseteq x \wedge \exists z\Big(z \subsetneq y \wedge \exists f\colon z \xrightarrow[\text{onto}]{\text{1-1}} y\Big)\Big)[a]$$

or $\mathfrak{A}_0 \models$ "x is infinite" $[a]$, as required.

PROOF OF LEMMA 2. Now A *is* infinite, so there is a $B \subsetneq A$ and a

$$G\colon B \xrightarrow[\text{onto}]{\text{1-1}} A.$$

All the formulas involved are setbounded in $\mathfrak{A}_0$ at $[G, B, A]$ and true in V at $[G, B, A]$, so by the proposition, the lemma is established.

This completes the proof of the theorem.

§VII. Some remarks on parameters. The language $\mathscr{L}$ considered so far contains only class terms of the form $\hat{x}\varphi$, where x is the only free variable in φ. φ can contain other class terms, but otherwise it involves only the binary predicate '$\in$'. Because definite descriptions cannot function in the usual way, the expressive power of φ is limited; for example, we cannot form something like $\hat{x}(x \in \{\varnothing\})$. This difficulty can be overcome by the introduction of set parameters into class terms. There are at least two ways to do this.

Begin by forming L_n, F_n and T_n as before, except this time let F_n contain all formulas of L_n, not just those with one free variable. Let $T = \bigcup_{n \in N} T_n$. Now we have class terms with free variables, such as $\hat{x}(x \in y)$ or $\hat{z}(\forall w(w \in z \equiv w \in y))$.

There are two ways to allow sets as parameters in place of these free variables. The easiest is to think of the class terms of the language as carrying their parameter assignments with them; that is, let

$$T_p = \{(t, \bar{p}) \mid t \in T \text{ and } \bar{p} \text{ is an assignment of members of } V \text{ to the free variables of } t\}$$

and let $\mathscr{L}_p$ be the first order language with the binary predicate '$\in$' and the members of T_p as constants. Then an $\mathscr{L}_p$-structure $\mathfrak{A}$ will be of the form $S_{\mathfrak{A}} \cup C_{\mathfrak{A}}$ where the members of $C_{\mathfrak{A}}$ are of the form $((t, \bar{p}), (t, \bar{p})^+, (t, \bar{p})^-)$ and there is one such element of $C_{\mathfrak{A}}$ for each $(t, \bar{p}) \in T_p$. $(t, \bar{p})^+$ and $(t, \bar{p})^-$ will be subcollections of $S_{\mathfrak{A}} \cup T_p$. Using these definitions, the construction can be carried out just as before, except that for $t = \hat{x}\varphi$,

$$(t, \bar{p})^+_{\alpha+1} = \{a \in V \mid \mathfrak{A}_\alpha \models \varphi[\bar{p}, a]\} \cup \{(u, \bar{q}) \in T_p \mid \mathfrak{A}_\alpha \models \varphi[\bar{q}, (u, \bar{q})^{\mathfrak{A}_\alpha}]\}$$

and similarly for $(t, \bar{p})^-$. (In $\varphi[\bar{p}, a]$, the $\bar{p}$ assignments inside t are to be made first. Strictly speaking, it would be best to switch to a new variable if x occurs free elsewhere in φ.) In this system, the natural numbers can be defined more or less as Frege intended.

DEFINITION.

$$0 = \hat{x}\left(\exists f\colon x \xrightarrow[\text{onto}]{\text{1-1}} \varnothing\right)\colon \quad 1 = \hat{x}\left(\exists f\colon x \xrightarrow[\text{onto}]{\text{1-1}} \{\varnothing\}\right), \quad \text{etc.}$$

The disadvantage of this straightforward method of introducing parameters is that it is impossible to quantify into class terms. For example, we cannot ask whether $\exists x(y \in \hat{z}(z \in x))$ is true at some y or other. To allow for this, the class terms cannot be thought of as carrying their parameters with them. Rather, the assignments to the free variables of members of T will have to be made by the $\mathfrak{A}$-sequences s. This is easily begun by forming the first order language $\mathscr{L}^*$ with the binary predicate '$\in$' and our new T for constants. $\mathscr{L}^*$-structures will contain interpretations for ordered pairs of the form $(t, \bar{p})$, where $t \in T$ and $\bar{p}$ is an assignment of members of $S_{\mathfrak{A}}$ to the free variables of t. The added complexity comes with the definition of an $\mathfrak{A}$-sequence s. It is not enough to say that $s(t)$ is $(t, \bar{p})$, where $\bar{p}$ is the assignment s makes to the free variables of t because s might assign members of $C_{\mathfrak{A}}$ to some of these variables, and our $\mathscr{L}^*$-structure $\mathfrak{A}$ only contains interpretations for $(t, \bar{p})$'s where the $\bar{p}$ assignments are in $S_{\mathfrak{A}}$. (The definition of $\mathfrak{A}$ becomes viciously circular otherwise.) This difficulty can be overcome as follows: if s assigns a class to some free variable of t, substitute the term for that class into t and extend s to assign the appropriate members of $S_{\mathfrak{A}}$ to the free variables of the new, more complicated term. This may require some switching of free variables. More precisely, if s assigns $(u, \bar{q})^{\mathfrak{A}}$ to some free variable x of t, replace the free variables of u by new ones $v_0, \ldots, v_n$ which do not occur in φ to form u', substitute u' for x in t to form t', evaluate t' at $s(v_0, \ldots, v_n/\bar{q})$. Then $\mathfrak{A} \models \varphi[s]$ can be defined as usual, and the construction carried out as before.

This second method of introducing parameters is only slightly more cumbersome, and it provides much more flexibility. For example, not only can the natural numbers be defined à la Frege, they can also be quantified over. "All natural numbers are φ" becomes

$$\forall z\left(z \in y \supset \varphi\left(\hat{x}\left(\exists f\colon x \xrightarrow[\text{onto}]{1\text{-}1} z\right)\right)\right)$$

where y is assigned the set of finite von Neumann ordinals. So, the second method is clearly preferable.

§VIII. Conclusion. I think this sketch goes some distance toward providing a theory of both sets and classes along Königean lines. The sets of the semantics are understood combinatorially, just as in Zermelo's iterative hierarchy, and the classes are understood logically, as extensions. (Though, as I have said, the restriction to terms of $\mathscr{L}$, or even of $\mathscr{L}_p$, is artificial.) Recalling the two desiderata at the end of §II, it seems clear that (2) is satisfied. Classes are significantly different from sets in several ways: they are logical, not combinatorial, they can be self-membered, and some membership relations involving them are indeterminate. The situation with respect to (1) is only slightly less clear. There is no trace in this account of the neo-Aristotelean ill-defined entities, and the indeterminacy of membership results from a type of ungroundedness, not from any sort of vagueness. On the other hand, since I have not presented a reduction of classes to other entities, I have not conclusively shown them to be "real". But I can conclude that unlike other theories which satisfy (2), mine is at least consistent with (1). Finally, I think that tracing the Fregean class paradoxes to the intuitively satisfying notion of ungroundedness allows us to say, with Gödel, that our logical intuitions are essentially correct, only somewhat blurred.

Let me conclude then by saying that I hope the construction presented here might provide the groundwork for an acceptable theory of sets and classes. Among the many questions that remain open are the following:

(1) Does the construction reach a fixed point? If so, where?

(2) Are there any t's such that $t^+ \cup t^- = V \cup T$? In particular, is there a t such that $t^+ = V$ and $t^- = T$?

(3) What happens to the construction if V is replaced by R_κ for some inaccessible κ?

(4) How and to what extent can the theory of V^* be axiomatized?

UNIVERSITY OF NOTRE DAME
NOTRE DAME, INDIANA 46556