Maddy's Theory of Classes
Alternative Set Theory Seminar

David Santamaría

February 19, 2018

In her search for a realistic theory of sets and classes, Maddy begins 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. Von Neumann, Morse and Kelley 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. On the other hand, concentrating on (2) leads to Cantor's nonactual, or ineffable proper classes, or the official Ackermann's ill-defined entities. 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).

Maddy wants 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, pretheoretically, some extensions seem to be members of others (e.g. $\widehat{x}$ (x is infinite)$\in \widehat{x}$(x has more than three elements)), and some extensions even appear to be self-membered (e.g., $\widehat{x}$(x is infinite) $\in \widehat{x}$ (x is infinite)). 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 $\widehat{x}(x \notin x)$? 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. But there is a third option. To see what it is, consider a similar situation, 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 Kripke has shown how the truth paradoxes can be solved by allowing truth value gaps as specified by a certain construction. What Maddy proposes 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. Maddy proposes 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.

We consider a FOL $\mathfrak{L}$ with $=$ and $\in$ as nonlogical symbols and we add a term-forming operator $\widehat{}$ to form terms such as $\widehat{x}(x = x)$ and $\widehat{x}(x \in \emptyset)$. To gain expressive power we will also include a constant $\overline{V}$ to stand for the class of all sets, and a constant $\overline{a}$.

Definition 1 (*Terms and formulas of* $\mathfrak{L}$)**.**

1. All constants and variables are terms.
2. If t and t' are terms, then $t = t'$ and $t \in t'$ are formulas.
3. If ϕ and ψ are formulas, and x is a variable, then $\neg\phi, \phi \wedge \psi$ and $\forall x \phi$ are formulas.
4. If ϕ is a formula, and x is among the free variables of ϕ, then $\widehat{x}\phi$ is a term.

T is the collection of all terms, it is the union of S, the collection of all set constants, C, the collection of all terms of the form $\widehat{x}\phi$, and $\{\overline{V}\}$. C^* is the collection of closed terms in C, similarly, T^* is the collection of closed terms in T.

The standard model for this language contains all sets, $\overline{a}$ standing for a, here $\overline{V}$ will be a class with extension all sets and antiextension all classes. The variable part of the interpretations of $\mathfrak{L}$ is the extension and antiextension of the elements of C^*.

Definition 2 ($\mathfrak{L} - structure$)**.** $\mathfrak{C} = \{(t, t^+_{\mathfrak{C}}, t^-_{\mathfrak{C}}) : t \in C^*\}$ is an $\mathfrak{L}$-structure iff $\forall t \in C^*$, $t^+_{\mathfrak{C}} \subseteq T^*$ and $t^-_{\mathfrak{C}} \subseteq T^*$ and $t^+_{\mathfrak{C}} \cap t^-_{\mathfrak{C}} = \emptyset$.

Note that it needn't be the case that $t^+_{\mathfrak{C}} \cup t^-_{\mathfrak{C}} = T^*$, and so we can have membership gaps. The idea is that $t^+_{\mathfrak{C}}$ and $t^-_{\mathfrak{C}} = \emptyset$ represent the extension and antiextension respectively of the class term t.

Given a sentence τ we have three possibilities $\mathfrak{C} \vDash \tau$ ($\mathfrak{C}$ thinks τ is true), $\mathfrak{C} \nvDash \tau$ ($\mathfrak{C}$ thinks τ is false) and $\mathfrak{C} \vDash^? \tau$ ($\mathfrak{C}$ does not have an opinion about τ).

Definition 3 (*Semantics for atomic sentences*)**.**

- τ is of the form $t \in t'$ for $t, t' \in T^*$, then:
  - $\mathfrak{C} \vDash \tau$ iff
    1. t is $\overline{a}$, t' is $\overline{b}$, and $a \in b$, or
    2. $t \in S$ and t' is $\overline{V}$, or

3. $t' \in \mathfrak{C}^*$ and $t \in (t')^+_{\mathfrak{C}}$

– $\mathfrak{C} \not\models \tau$ iff

1. t is $\overline{a}$, t' is $\overline{b}$, and $a \notin b$, or
2. t is $\overline{V}$ and $t' \in S \cup \{\overline{V}\}$, or
3. $t \in \mathfrak{C}^*$ and $t' \in S \cup \{\overline{V}$, or
4. $t' \in \mathfrak{C}^*$ and $t \in (t')^-_{\mathfrak{C}}$

- If τ is of the form $t = t'$ for $t, t' \in T^*$, then $\mathfrak{C} \models \tau$ iff t and t' are the same term, and $\mathfrak{C} \not\models \tau$ iff t and t' are different terms.

For complex sentences truth and falsity is defined via the strong Kleene rules.

Definition 4 (*Semantics for complex sentences*)**.** For sentences σ and τ,

1. $\mathfrak{C} \models \neg\sigma$ iff $\mathfrak{C} \not\models \sigma$; $\mathfrak{C} \not\models \neg\sigma$ iff $\mathfrak{C} \models \sigma$.
2. $\mathfrak{C} \models \sigma \wedge \tau$ iff $\mathfrak{C} \models \sigma$ and $\mathfrak{C} \models \tau$; $\mathfrak{C} \not\models \sigma \wedge \tau$ iff $\mathfrak{C} \not\models \sigma$ or $\mathfrak{C} \not\models \tau$.
3. $\mathfrak{C} \models \forall x\phi$ iff for all $t \in T^*, \mathfrak{C} \models \phi(t/x)$; $\mathfrak{C} \not\models \forall x\phi$ iff for some $t \in T^*, \mathfrak{C} \not\models \phi(t/x)$.

One can then define $\vee, \rightarrow, \equiv$ and $\exists$ from these in the usual way.

Definition 5. If $\mathfrak{C}$ and $\mathfrak{C}'$ are two $\mathfrak{L}-$structures, then $\mathfrak{C} \sqsubseteq \mathfrak{C}'$ iff for all $t \in C^*, t^+_{\mathfrak{C}} \subseteq t^+_{\mathfrak{C}'}$ and $t^-_{\mathfrak{C}} \subseteq t^-_{\mathfrak{C}'}$

The following result can be shown by induction on the complexity of formula:

Proposition 1 (Monotonicity)**.** *If $\mathfrak{C} \sqsubseteq \mathfrak{C}'$, then for any sentence σ, if $\mathfrak{C} \models \sigma$, then $\mathfrak{C}' \models \sigma$, and if $\mathfrak{C} \not\models \sigma$, then $\mathfrak{C}' \not\models \sigma$.*

This tells us that once a sentence is decided, adding more elements to the extensions and antiextensions of classes does not disturb this fact.

With this machinery in place we construct the following sequence of $\mathfrak{L}-$structures:

$$\mathfrak{C}_0 = \{(\widehat{x}\phi, \widehat{x}\phi^+_0, \widehat{x}\phi^-_0) : \widehat{x}\phi \in C^*\} \text{ where } \widehat{x}\phi^+_0 = \widehat{x}\phi^-_0 = \emptyset$$

$$\mathfrak{C}_{\alpha+1} = \{(\widehat{x}\phi, \widehat{x}\phi^+_{\alpha+1}, \widehat{x}\phi^-_{\alpha+1}) : \widehat{x}\phi \in C^*\} \text{ where } \begin{cases} \widehat{x}\phi^+_{\alpha+1} = \{t \in T^* : \mathfrak{C}_\alpha \models \phi(t/x)\} \\ \widehat{x}\phi^-_{\alpha+1} = \{t \in T^* : \mathfrak{C}_\alpha \not\models \phi(t/x)\} \end{cases}$$

For λ a limit ordinal,

$$\mathfrak{C}_\lambda = \{(\widehat{x}\phi, \widehat{x}\phi^+_\lambda, \widehat{x}\phi^-_\lambda) : \widehat{x}\phi \in C^*\} \text{ where } \begin{cases} \widehat{x}\phi^+_\lambda = \bigcup_{\alpha<\lambda} \widehat{x}\phi^+_\alpha \\ \widehat{x}\phi^-_\lambda = \bigcup_{\alpha<\lambda} \widehat{x}\phi^-_\alpha \end{cases}$$

Define an $\mathfrak{L}-$structure U (for universe):

$$U = \{(\widehat{x}\phi, \widehat{x}\phi^+, \widehat{x}\phi^-) : \widehat{x}\phi \in C^*\} \text{ where } \begin{cases} \widehat{x}\phi^+ = \bigcup_{\alpha\in Ord} \widehat{x}\phi^+_\alpha \\ \widehat{x}\phi^- = \bigcup_{\alpha\in Ord} \widehat{x}\phi^-_\alpha \end{cases}$$

We are interested in U. By monotonicity, whatever becomes true or false at one of the $\mathfrak{C}_\alpha$, remains true or false in U. For example, $\mathfrak{C}_0 \models \overline{\emptyset} \in \overline{\{\emptyset\}}$, so $\overline{\emptyset} \in \widehat{x}(x \in \overline{\{\emptyset\}})^+_1$, thus $\mathfrak{C}_1 \models \overline{\emptyset} \in \widehat{x}(x \in \overline{\{\emptyset\}})$ and so $U \models \overline{\emptyset} \in \widehat{x}(x \in \overline{\{\emptyset\}})$.

Definition 6 (*Ordered tuple*)**.** For $t, t' \in T^*$, '(t, t')' is $\widehat{z}(z = t \vee z = t')$.

Note that since $\widehat{z}(z = t \vee z = t')$ is not the same symbol as $\widehat{z}(z = t' \vee z = t)$ it is easily established that:

Proposition 2 (Equality of ordered tuples)**.** *For* $t, t' \in T^*, U \models ((t, t') = (u, u'))$ *iff* $U \models (t = u \wedge t' = u')$.

Notice also that these ordered classes are total, and so $U \not\models ((t, t') = (u, u'))$ iff $U \not\models (t = u \wedge t' = u')$. We can define ordered n-tuples as usual: $(t, t', t'') = ((t, t'), t'')$. Now, if $x_0, \ldots, x_n$ are among the free variables of ϕ then $\widehat{x}_0, \ldots, \widehat{x}_n\phi$ abbreviates $\widehat{z}(\exists x_0, \ldots, \exists x_n(z = (x_0, \ldots, x_n) \wedge \phi))$, with z, the first variable not appearing in ϕ.

Continuing with the last example, recall that $\mathfrak{C}_0 \models \overline{\emptyset} \in \overline{\{\emptyset\}}$ and so $\mathfrak{C}_0 \models \exists x \exists y((\overline{\emptyset} \in \overline{\{\emptyset\}}) = (x, y) \wedge x \in y)$, which means $(\overline{\emptyset}, \overline{\{\emptyset\}}) \in \widehat{z}(\exists x \exists y(z = (x, y) \wedge x \in y))^+_1$ and so $\mathfrak{C}_1 \models ((\overline{\emptyset}, \overline{\{\emptyset\}}) \in \widehat{x}\widehat{y}(x \in y))$ and by monotonicity, U agrees.

Now we can prove what Maddy considers one of the great advantages of her system:

Theorem 1. $U \models \widehat{x}(x \text{ is infinite}) \in \widehat{x}(x \text{ is infinite})$

Proof. We say that a collection, x, is relational iff $\forall y(y \in x \rightarrow \exists u \exists v(y = (u, v)))$. A non-empty relational collection will be a class, not a set, because it has classes as members. We also define a domain and a range of a relational class as usual, same for functional and one-to-one. For any set function f, let f^* be $\widehat{x}(\exists y \exists z(\langle y, z\rangle \in \overline{f} \wedge x = (y, z)))$, where $\langle y, z\rangle \in \overline{f}$ is the usual statement that the Kuratowski ordered pair of y and z is in $\overline{f}$, but with y, z and all quantifiers relativised to $\overline{V}$. Then one can prove that for sets a and b:

1. If $\langle y, z\rangle \in f$, then $\mathfrak{C}_1 \models (\overline{a}, \overline{b}) \in f^*$.
2. If $\langle y, z\rangle \notin f$, then $\mathfrak{C}_1 \not\models (\overline{a}, \overline{b}) \in f^*$.
3. $\mathfrak{C}_1 \models \forall x(x \in f^* \vee x \notin f^*)$ (i.e. f^* is total).

So we see that f^* will behave as a class surrogate of f in $\mathfrak{C}_1$.

Definition 7 (Infinity)**.** 'x is infinite' abbreviates $\exists f($ 'f is functional' $\wedge$ 'f is one-to-one'$\wedge$ '$\overline{\omega}$ is a domain of f' $\wedge$ 'x contains a range of f').

For any $n \in \omega$ let $n^* = \{n+1, n+2, \ldots\}$. This set is clearly infinite; using f^*, where f is a set function that maps ω one-to-one into n^*, we can also show that for all $n \in \omega, \mathfrak{C}_2 \models \overline{n^*} \in \widehat{x}(x \text{ is infinite})$. Now let $\phi(y, z)$ be the formula $y \in \overline{\omega} \wedge z \in \overline{\mathcal{P}(\omega)} \wedge \forall u(u \in \overline{V} \rightarrow (u \in z \equiv u \in \overline{\omega} \wedge y \in u))$. Then, for all

$n \in \omega, \mathfrak{C}_1 \vDash (\overline{n}, \overline{n^*}) \in \widehat{y}\widehat{z}\phi$. $\mathfrak{C}_1$ also thinks that $\widehat{y}\widehat{z}\phi$ is functional and one-to-one, and that ω is a domain, by the preceding result, $\mathfrak{C}_2$ thinks that $\widehat{x}(x$ is infinite) contains a range. The result follows. Q.E.D.

Finally, the gaps in the membership relation allow for the Russell paradox to be sidestepped:

Theorem 2. $U \vDash^? \widehat{x}(x \notin x) \in \widehat{x}(x \notin x)$

Proof. If $\widehat{x}(x \notin x)$ were in $[\widehat{x}(x \notin x)]^+$, then it would have to enter at some $\widehat{x}(x \notin x)^+_\alpha$. Because α cannot be a limit, it must be of the form $\beta + 1$. But then, $\mathfrak{C}_\beta \vDash \widehat{x}(x \notin x) \notin \widehat{x}(x \notin x)$, and so $\mathfrak{C}_\alpha \vDash \widehat{x}(x \notin x) \in \widehat{x}(x \notin x)$ (because $\widehat{x}(x \notin x) \in \widehat{x}(x \notin x)^+_\alpha$) and $\mathfrak{C}_\alpha \vDash \widehat{x}(x \notin x) \notin \widehat{x}(x \notin x)$ (because $\mathfrak{C}_\beta \sqsubseteq \mathfrak{C}_\alpha$). Contradiction. Similarly, $\widehat{x}(x \notin x) \notin [\widehat{x}(x \notin x)]^-$. Q.E.D.

Further Reading:

Maddy, Penelope:

[1983] *Proper classes*, Journal of Symbolic Logic 48, pp. 113-139.

[2000] *A theory of classes*, in G. Sher and R. Tieszen, eds., *Between Logic and Intuition: Essays in Honor of Charles Parsons*, (Cambridge: Cambridge University Press), pp. 299 – 316.

Appendix: Talk Handout

*Text 1:

. . . 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.

+Stifel, Quoted in M. Kline, *Mathematical thought from ancient to modern times,* Oxford, 1972, p. 251

*Text 2:

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

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."

*Text 3:

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.

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

*Text 4:

(a)

. . . 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' . . . .

(b)

. . . 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' . . .

+Cantor, *Letter to Dedekind,* in *Mathematical logic from Frege to Gödel* (J. van Heijenoort, ed.), Harvard University Press, Cambridge, Mass., 1967, p. 114

*Text 5:

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".

+ 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.

*Text 6:

When the notion of the continuum is formed, it is the 'arbitrary' sequence (a_1, a_2, ..., a_k, ...) that is primary, or fundamental. Through the stipulation that a_1, a_2, ... 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 [ω_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 [ω_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, *On the foundations of set theory and the continuum problem* in van Heijenoort, op. cit., pp. 145-149.

*Text 7:

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.

+ Bernays, *On Platonism in mathematics* in P. Benacerraf and H. Putnam, *Philosophy of mathematics*, Prentice-Hall, Princeton, N.J., 1964 pp. 275-276.

*Text 8:

... 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.

+ D. Martin, Sets versus classes, circulated xerox.

*Text 9:

... 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.

+ Gödel, *What is Cantor's continuum problem?,* pp. 262-263.

*Text 10:

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.

+ Gödel, *Russell's mathematical logic*, p. 229.