

David Ellerman

CATEGORY THEORY AND SET THEORY AS THEORIES ABOUT COMPLEMENTARY TYPES OF UNIVERSALS

Abstract. Instead of the half-century old foundational feud between set theory and category theory, this paper argues that they are theories about two different complementary types of universals. The set-theoretic antinomies forced naïve set theory to be reformulated using some iterative notion of a set so that a set would always have higher type or rank than its members. Then the universal $u_F = \{x \mid F(x)\}$ for a property $F(\cdot)$ could never be self-predicative in the sense of $u_F \in u_F$. But the mathematical theory of categories, dating from the mid-twentieth century, includes a theory of always-self-predicative universals – which can be seen as forming the “other bookend” to the never-self-predicative universals of set theory. The self-predicative universals of category theory show that the problem in the antinomies was not self-predication *per se*, but negated self-predication. They also provide a model (in the Platonic Heaven of mathematics) for the self-predicative strand of Plato’s Theory of Forms as well as for the idea of a “concrete universal” in Hegel and similar ideas of paradigmatic exemplars in ordinary thought.

Keywords: universals; category theory; Plato’s Theory of Forms; set theoretic antinomies; universal mapping properties

1. Introduction

Over the last half century, there has been a foundational feud between set theory and category theory. The purpose of this paper is to argue that this (often silly) feud should be replaced by the view that, at least as

theories of universals, the two theories are quite complementary book-ends: set theory as the theory of never-self-predicative universals and category theory as the theory of always self-predicative universals.

According to the naïve concept of a set, given a predicate F (not involving y), then there exists a set y such that for all x , $x \in y$ if and only if (iff) $F(x)$ [4, p. 217]. Such a comprehension scheme establishes y as a type of universal, which might be denoted as u_F , for the property $F(\cdot)$. But the set-theoretic antinomies forced the naïve set theory of Frege’s Paradise to be reconstructed in accordance with models designed to prevent any self-membership (see [4, 31]).¹ In the formal systems of axiomatic set theory, self-membership $u_F \in u_F$ (e.g., for the property $F(x) \equiv x \notin x$) was ruled out, e.g., by the axiom of regularity in Zermelo-Fraenkel set theory. The set-universals in the resulting models were always non-self-predicative. Self-predication was even seen as the culprit in the antinomies.

The idea of universals goes back to Plato’s theory of ideas (the epsilon of set membership derives from $\varepsilon\iota\delta\eta$ for ideas), and the self-predicative version sounds, to modern ears, like some “bad Platonic metaphysics” such as the following:

Given all the entities that have a certain property, there is one entity among them that exemplifies the property in a universal, paradigmatic, archetypical, ideal, essential, or canonical way. It is called the “self-predicative universal.” There is a relationship of “participation” or “resemblance” so that all and only the entities that have the property (including itself) “participate in” or “resemble” that perfect example, the self-predicative universal.

But self-predication has taken a “bum rap.” Around the middle of the 20th century, a new mathematical theory, category theory (Mac Lane [18]; Awodey [3]), was developed which featured universals that were always self-predicative. The antinomies showed there could not be *one* theory where the universals could be non-self-predicative and also self-predicative. Instead, there are two theories of universals, one theory where the set-universals are never self-predicative (with set membership

¹ The rough idea is to start off some *urelements* (individuals not involving sets) or possibly none, and at stage zero, construct all sets of *urelements* (including the empty set). At stage one, construct all sets of sets of *urelements* and of *urelements*. At each stage, a set may contain any *urelements* or sets constructed at an earlier stage—so no set could be a member of itself.

as the participation relation) as well as the other bookend where the universals are always self-predicative (the universal morphisms of category theory).

A simple example using sets will illustrate the points. Given two sets a and b , consider the property of sets: $F(x) \equiv x \subseteq a \ \& \ x \subseteq b$. In other words, the property is the property of being both a subset of a and a subset of b . In this example, the participation relation is the subset *inclusion* relation (which, unlike set membership, is a morphism in the category theoretic sense). There is a set, namely the intersection or meet of a and b , denoted $a \cap b$, that has the property (so it is a “concrete” instance of the property), and it is universal in the sense that any other set has the property if and only if it participates in (i.e., is included in) the universal example:

self-predication: $F(a \cap b)$, i.e., $a \cap b \subseteq a$ and $a \cap b \subseteq b$, and

universality: x participates in $a \cap b$ if and only if $F(x)$, i.e., $x \subseteq a \cap b$ if and only if $x \subseteq a$ and $x \subseteq b$.

This example of a self-predicative universal is quite simple, but all this “bad metaphysical talk” has highly developed and precise models in category theory.

In terms of the old theme of universals in philosophy, the self-predicative universals of category theory provide a rigorous model (in the “Platonic Heaven” of mathematics) for the self-predicative strand in Plato’s thought and the “concrete universal” (Ellerman [6]; Honderich [12, Entry at “Concrete Universal”]) version of synthesis in Hegel’s thought as well as for the common “Form” of thought that considers a paradigmatic (Sayre [30]), canonical, iconic, archetypical, or quintessential exemplar of some property. Perhaps due to their mathematical nature, these self-predicative universals of category theory seem to be unknown in the contemporary philosophical literature (e.g., [8]).

2. Criteria for a Theory of Universals

In Plato’s Theory of Ideas or Forms ($\epsilon\iota\delta\eta$), a property F has an entity associated with it, the universal u_F , which uniquely represents the property. An object x has the property F , i.e., $F(x)$, if and only if (iff) the object x participates in the universal u_F .

Let μ (from $\mu\epsilon\theta\epsilon\xi\iota\varsigma$ or *methexis* for participation) represent the participation relation so

‘ $x \mu u_F$ ’ reads as “ x participates in u_F ”.

Given a relation μ , an entity u_F is said to be a *universal* for the property F (with respect to μ) if it satisfies the following universality condition:

for any x , $x \mu u_F$ if and only if $F(x)$.

A universal representing a property should be in some sense unique. Hence there should be an equivalence relation (\approx) so that universals satisfy a uniqueness condition:

if u_F and u'_F are universals for the same F , then $u_F \approx u'_F$.

A mathematical theory is said to be a *theory of universals* [6] if it contains a binary relation μ and an equivalence relation \approx so that with certain properties F there are associated entities u_F satisfying the following conditions:

- (I) Universality: for any x , $x \mu u_F$ iff $F(x)$, and
- (II) Uniqueness: if u_F and u'_F are universals for the same F , then $u_F \approx u'_F$.

A universal u_F is said to be *abstract* or *non-self-predicative* if it does not participate in itself, i.e., u_F). A universal u_F is *self-predicative* or *concrete* if it participates in itself, i.e., $u_F \mu u_F$.

3. Set Theory as the Theory of Non-self-predicative Universals

There is a modern mathematical theory that readily qualifies as a theory of universals, namely set theory. In the naïve form of set theory (“Frege’s Paradise”), the universal representing a property F is the set of all elements with the property:

$$u_F = \{x \mid F(x)\}.$$

The participation relation is the set membership relation. The universality condition in (naïve) set theory is the comprehension axiom: there is a set y such that for any x , $x \in y$ iff $F(x)$. Set theory also has an extensionality axiom, which states that two sets with the same members are identical:

for all x , $(x \in y \text{ iff } x \in y')$ implies $y = y'$.

Thus, if y and y' both satisfy the comprehension axiom scheme for the same F then y and y' have the same members so $y = y'$. Hence in set theory, the uniqueness condition on universals is satisfied with the equivalence relation (\approx) as equality ($=$) between sets. Thus naïve set theory satisfies the conditions for a theory of universals – if only it was consistent. The naïve comprehension axiom lead to inconsistency for such properties as

$$F(x) \equiv \text{“}x \text{ is not a member of } x\text{”} \equiv x \notin x.$$

If R is the universal for that property, i.e., R is the set of all sets which are not members of themselves, the naïve comprehension axiom yields a contradiction (*Russell’s Paradox*):

$$R \in R \text{ iff } R \notin R.$$

The characteristic feature of Russell’s Paradox and the other set theoretical paradoxes is the negated self-reference wherein the universal is allowed to qualify for the negated property represented by the universal, e.g., the Russell set R is allowed to be one of the x ’s in the universality relation: $x \in R$ iff $x \notin x$. The set-theoretic formulation of the paradox was not essential. Russell himself expounded the paradox in term of the property of predicates that they are not self-predicative, i.e., the “predicates which are not predicable of themselves” (Russell [28, p. 80]).

There are several ways to restrict the naïve comprehension axiom to defeat the set theoretical paradoxes, e.g., as in Russell’s type theory, Zermelo-Fraenkel set theory, or von-Neumann-Bernays set theory. The various restrictions are based on an iterative concept of set (Boolos [4]) which forces a set y to be more “abstract”, e.g., of higher type or rank (e.g., [31, p. 304]), than the elements $x \in y$. As Russell himself put it using a notion of “degree”:

It will now be necessary to distinguish (1) terms, (2) classes, (3) classes of classes, and so on ad infinitum; we shall have to hold that no member of one set is a member of any other set, and that $x \in u$ requires that x should be of a set of a degree lower by one than the set to which u belongs. Thus $x \in x$ will become a meaningless proposition; and in this way the contradiction is avoided. [28, p. 527]

Thus the universals provided by the various set theories are “abstract” universals in the technical sense that they are relatively more ab-

stract (i.e., of higher type or rank) than the objects having the property represented by the universal. Sets may not be members of themselves.²

With the modifications to avoid the paradoxes, a set theory still qualifies as a theory of universals. For suitably restricted predicates, there exists a set of the elements with that property and the set satisfies the universality condition with membership as the participation relation and extensional equality as the equivalence relation. But set theory cannot qualify as a general theory of universals. The paradox-induced modifications turn the various set theories into theories of abstract (never-self-predicative) universals since they prohibit the self-membership of sets. That clears the ground for another theory of always-self-predicative universals.

4. Self-Predicative or Concrete Universals in Philosophy

Philosophy has long contemplated another type of universal, variously called a *self-predicative*, *self-participating*, or *concrete universal*. Indeed, it is a common Form of thought. The intuitive idea of a self-participating universal for a property is that it is an object that has the property and has it in such a universal sense that all other objects with the property resemble or participate in that paradigmatic, archetypal, canonical, iconic, ideal, essential, or quintessential exemplar. Such a universal u_F for a property F is self-predicative in the sense that it has the property itself, i.e., $F(u_F)$. It is universal in the intuitive sense that it represents F -ness in such a perfect and exemplary manner that any object resembles or participates in the universal u_F if and only if it has the property F .

If anything else is beautiful besides Beauty itself, it is beautiful for no other reason than because it participates in that Beauty.

Phaedo 100, translation from [7, p. 35]

The intuitive notion of a concrete universal or paradigmatic instance occurs in ordinary thought as in the “all-American boy” or any case of a quintessential iconic example (e.g., Sophia Loren as “the” Italian women or Michelangelo’s David or da Vinci’s Mona Lisa as “the” exemplars for certain artistic categories), or when resemblance to an “defining” example becomes an adjective like “Lincolnesque.” In Greek-inspired Christian

² Quine’s system ML [27] allows “ $V \in V$ ” for the universal class V , but no standard model of ML has ever been found where “ \in ” is interpreted as set membership (Hatcher [11, Chapter 7]).

theology, there is the “Word made Flesh” (Miles [24]) together with *imitatio Christi* as the participation or resemblance relation to that concrete universal. The idea of the concrete universal is often associated with Hegel (Stern [32]) where it was one way to think about the synthesis between an abstract universal thesis and the antithesis of diverse particulars. One sensible application by Hegel was in the arts and literature (Wimsatt [37]) to explicate the old idea that great art uses a concrete instance to universally exemplify certain human conditions, e.g., Shakespeare’s *Romeo and Juliet* as “the” romantic tragedy (Desmond [5]).

The notion of a self-predicative universal goes back to Plato’s Theory of Forms (Vlastos [33, 34, 35]; Malcolm [22]). Plato’s forms are often considered to be abstract or non-self-predicative universals quite distinct from and “above” the instances. In the words of one Plato scholar, “not even God can scratch Doghood behind the Ears” (Allen [1]). But Plato did give examples of self-participation or self-predication, e.g., that Justice is just [Protagoras 330] or that Beauty is beautiful. Moreover, Plato used expressions that indicated self-predication of universals.

But Plato also used language which suggests not only that the Forms exist separately ($\chi\omega\rho\iota\sigma\tau\alpha$) from all the particulars, but also that each Form is a peculiarly accurate or good particular of its own kind, i.e., the standard particular of the kind in question or the model ($\pi\alpha\rho\alpha\delta\epsilon\iota\gamma\mu\alpha$) to which other particulars approximate. [14, p. 19]

But many scholars regard the notion of a Form as *paradeigma* or self-predicative universal as an error.

For general characters are not characterized by themselves: humanity is not human. The mistake is encouraged by the fact that in Greek the same phrase may signify both the concrete and the abstract, e.g. $\lambda\epsilon\upsilon\kappa\omicron\nu$ (literally “the white”) both “the white thing” and “whiteness”, so that it is doubtful whether $\alpha\upsilon\tau\omicron\tau\omicron\lambda\epsilon\upsilon\kappa\omicron\nu$ (literally “the white itself”) means “the superlatively white thing” or “whiteness in abstraction”. [14, pp. 19–20]

Thus, some Platonic language is ambivalent between interpreting a form as a paradigmatic exemplar (“the superlatively white thing”) and an abstract universal (“whiteness in abstraction”).

The literature on Plato has reached no resolution on the question of self-predication. Scholarship has left Plato on both sides of the fence; many universals are not self-predicative but some are. It is fitting that

Plato should exhibit this ambivalence since the self-predication issue has only come to a head in the 20th century with the set theoretical antinomies. Set theory had to be reconstructed as a theory of universals (for the set-membership relation) that were never self-predicative. That cleared the ground for a separate theory of universals that are always self-predicative. Such a theory of self-predicative universals would realize the self-predicative strand of Plato's Theory of Forms.

5. Self-predicative Universals in Partial Orders

A theory of (always) self-predicative universals would have an appropriate participation relation μ so that for certain properties F , there are entities u_F satisfying the universality condition:

$$\text{for any } x, x \mu u_F \text{ iff } F(x).$$

The universality condition and $F(u_F)$ imply that u_F is a concrete universal in the previously defined sense of being self-predicative, $u_F \mu u_F$. A theory of self-predicative universals would also have to have an equivalence relation so the self-predicative universals for the same property would be the universal up to that equivalence relation.

Is there a precise mathematical theory of the common Form of thought, the self-predicative universal? Our claim is that category theory is precisely that theory where the self-predicative universal objects are characterized by "universal mapping properties" (UMPs). Most (but not all) UMPs are expressed using adjoint functors because adjointness isolates "what is universal in mathematics" (Lawvere [15, p. 281])

The simplest examples of categories are partially ordered sets.³ Consider the universe of subsets or power set $\wp(U)$ of a set U with the inclusion relation \subseteq as the partial ordering relation. Given subsets a and b , consider the property

$$G(x) \equiv a \subseteq x \ \& \ b \subseteq x.$$

The participation relation is the reverse of set inclusion \supseteq and the union $a \cup b$ is the universal u_G for this property $G(x)$. The universality

³ A binary relation \leq on U is a *partial order* if for all $u, u', u'' \in U$, it is reflexive ($u \leq u$), transitive ($u \leq u'$ and $u' \leq u''$ imply $u \leq u''$), and anti-symmetric ($u \leq u'$ and $u' \leq u$ imply $u = u'$).

relation states that the union is the least upper bound of a and b in the inclusion ordering:

$$\text{for any } x, x \supseteq a \cup b \text{ iff } a \subseteq x \ \& \ b \subseteq x.$$

The universal has the property it represents, i.e., $a \subseteq a \cup b$ and $b \subseteq a \cup b$, so it is a self-predicative or concrete universal.⁴ Two self-predicative universals for the same property must participate in each other. In partially ordered sets, the antisymmetry condition, $y \subseteq y' \ \& \ y' \subseteq y$ implies $y = y'$, means that equality can serve as the equivalence relation in the uniqueness condition for universals in a partial order.

There is much controversy in Platonic scholarship about self-predication (see Malcolm [22] for a summary and more references). Our purpose is *not* Plato exegesis. But the development of a mathematical theory of self-predicative universals in category theory does help by showing one way to sort out Plato's Ideas.

One simple point is that there can be both non-self-predicative and self-predicative universals for the same property $F(x)$, and both are distinct from that property. Given subsets $a, b \subseteq U$, consider the property: $F(x) \equiv x \subseteq a \ \& \ x \subseteq b$. The self-predicative universal (or "paradigmatic instance" in Platonic language) for that property is the intersection $a \cap b$ where the participation relation is inclusion \subseteq (*Universality condition for self-predicative universal*):

$$\forall x \subseteq U: x \subseteq a \cap b \text{ iff } F(x)$$

The power set $\wp(a \cap b)$ is the non-self-predicative universal for that property where the participation relation is set membership (*Universality condition for non-self-predicative universal*):

$$\forall x \subseteq U: x \in \wp(a \cap b) \text{ iff } F(x).$$

The self-predicative universal $a \cap b$ has the property, i.e., $a \cap b \subseteq a \ \& \ a \cap b \subseteq b$, while the non-self-predicative universal $\wp(a \cap b)$ does not have the property, and neither universal is to be confused with the property $F(x) \equiv x \subseteq a \ \& \ x \subseteq b$ itself.

⁴ These universals are "concrete" in the technical sense of being "One among the many." They are not relatively abstract or "One over the many" like a universal in set theory relative to its members. Of course, the concrete universals of category theory are abstract (in the usual sense) mathematical entities in "Plato's Heaven" and are not ordinary concrete objects [23].

6. Self-predicative Universals in General Categories

The self-predicative universals of category theory⁵ are usually presented as universal mapping properties and the participation relation is the *uniquely-factors-through* relation. It can always be formulated in a suitable category as:

“ $x \mu u_F$ ” means “there exists a unique arrow $x \rightarrow u_F$ ”.

Then x is said to *uniquely factor through* u_F , and the arrow $x \rightarrow u_F$ is the unique factor or participation morphism. In the universality condition,

$$x \mu u_F \text{ if and only if } F(x),$$

the existence of the identity arrow $1_{u_F} : u_F \rightarrow u_F$ is the self-participation of the self-predicative universal that corresponds with $F(u_F)$, the self-predication of the property to u_F . In category theory, the equivalence relation used in the uniqueness condition is the isomorphism (\cong).⁶

It is sometimes convenient to “turn the arrows around” and use the dual definition where “ $x \mu u_G$ ” means “there exists a unique arrow $u_G \rightarrow x$ that can also be viewed as the original definition stated in the dual or opposite category. The above treatment of the intersection $a \cap b$ and the union $a \cup b$ are dual to one another. If we think of “participation” as *transferring* the property from the universal to the instances, then the transferring may go along the direction of the arrow — so the property may be said to be *transmitted* to the instance — or the transferring may go against the direction of the arrow — so the property may be said to be *reflected* to the instance. If we take the inclusion \subseteq as the direction of the arrow \rightarrow for a partial order as a category, then in the case of the union, $x \supseteq a \cup b$ iff $G(x)$, the union $a \cup b$ transmitted the property G to the particular x , and in the case of the intersection, $x \subseteq a \cap b$ iff $F(x)$, the intersection $a \cap b$ reflected the property F to the particular x .

⁵ For more philosophically-oriented introductions to category theory, see Lawvere and Schanuel [16], Magnan and Reyes [21], or Awodey [3].

⁶ Thus it must be verified that two concrete universals for the same property are isomorphic. By the universality condition, two concrete universals u and u' for the same property must participate in each other. Let $f : u' \rightarrow u$ and $g : u \rightarrow u'$ be the unique arrows given by the mutual participation. Then by composition $gf : u' \rightarrow u'$ is the unique arrow $u' \rightarrow u'$ but $1_{u'}$ is another such arrow so by uniqueness, $gf = 1_{u'}$. Similarly, $fg : u \rightarrow u$ is the unique self-participation arrow for u so $fg = 1_u$. Thus mutual participation of u and u' implies the isomorphism $u \cong u'$.

One of the earliest examples of a universal mapping property to be isolated and characterized was the direct product of, say, two sets A and B [17]. The property in question is:

$P(f, g) \equiv$ “ $f: X \rightarrow A$ and $g: X \rightarrow B$ is a pair or ‘cone’ $\{f, g\}$ of set functions from any set X to A and to B .”

The self-predicative universal for this property is the *pair of projection maps* $p_A: A \times B \rightarrow A$ and $p_B: A \times B \rightarrow B$ from the Cartesian product $A \times B = \{(a, b) : a \in A; b \in B\}$ to A and to B where $p_A((a, b)) = a$ and similarly for p_B . Then any such pair of maps $f: X \rightarrow A$ and $g: X \rightarrow B$ uniquely factors through the projections by the participation map $\langle f, g \rangle: X \rightarrow A \times B$ which takes $x \in X$ to $(f(x), g(x)) \in A \times B$. Then $\langle f, g \rangle$ followed by p_A gives f , i.e., $X \xrightarrow{\langle f, g \rangle} A \times B \xrightarrow{p_A} A = X \xrightarrow{f} A$, and $\langle f, g \rangle$ followed by p_B gives g , i.e., $X \xrightarrow{\langle f, g \rangle} A \times B \xrightarrow{p_B} B = X \xrightarrow{g} B$. That is, in the diagrammatic language of category theory, the participation map $\langle f, g \rangle$ is the unique map that makes the two triangles in the following diagram *commute*.

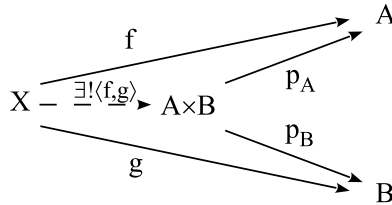


Figure 1. Universal mapping property for Cartesian product of sets

In this case, since the direction of the participation map $\langle f, g \rangle$ is towards the universal, we would say the universal pair, $\{p_A, p_B\}$ *reflects* the property back to the instance $\{f, g\}$ of the property. The universality condition is:

$\{f, g\}$ uniquely factors through $\{p_A, p_B\}$ iff $P(f, g)$.

The self-predicative universal $\{p_A, p_B\}$ uniquely factors through itself by the identity map $1_{A \times B}$ and the product $A \times B$ is unique up to isomorphism (in the category of sets and functions).

Another early example of a universal mapping property (Samuel [29]) was the free group $F[X]$ generated by a set X . The property in this case is:

$P(\varphi) \equiv$ “ $\varphi: X \rightarrow G$ is a map from the set X to any group G ”.

The self-predicative universal for this property is the *injection of the generators map* $\eta_X : X \rightarrow F[X]$. Then for any set-to-group map $\varphi : X \rightarrow G$ from the given set X to any group G , there is a unique group homomorphism $f : F[X] \rightarrow G$ so that the insertion of the generators η_X followed by f gives φ , i.e., $X \xrightarrow{\eta_X} F[X] \xrightarrow{f} G = X \xrightarrow{\varphi} G$. Hence f is the unique group homomorphism that makes the following diagram commute.

$$\begin{array}{ccc}
 X & & \\
 \eta_X \downarrow & \searrow \varphi & \\
 F[X] & \xrightarrow{\exists! f} & G
 \end{array}$$

Figure 2. Universal mapping property for the free group

In this case, the participation morphism f takes the universal η_X to the instance φ of the property so it *transmits* the property in question from the universal to the instance. The universality condition is:

$$\varphi \text{ uniquely factors through } \eta_X \text{ iff } P(\varphi).$$

This universality condition is just a reformulation of the usual universal mapping property of free groups. For instance, here is that UMP as stated in an algebra text (using our notation):

Let $F[X]$ be the free group on a set X and $\eta_X : X \rightarrow F[X]$ the inclusion map. If G is a group and $\varphi : X \rightarrow G$ a map of sets, then there exists a unique homomorphism of groups $f : F[X] \rightarrow G$ such that $f\eta_X = \varphi$.

[13, p. 65]

The self-predicative universal η_X uniquely factors through itself by the identity group homomorphism $1_{F[X]}$ and the free group $F[X]$ is unique up to isomorphism (in the category of groups and group homomorphisms). All the UMPs of category theory can be reformulated in a similar manner as self-predicative universals for certain mapping properties.

Category theory as the theory of self-predicative universals has quite a different flavor from set theory, the theory of abstract non-self-predicative universals. Given an appropriately delimited collection of all the elements with a property, set theory can postulate a more abstract entity, the set of those elements, to be the universal. But category theory cannot postulate its universals because those universals are self-predicative, i.e., are the “One among the many” — if any. Category theory must find its universals, if at all, among the many entities with the property.

7. Self-Predicative Universals and the Antinomies

The flaw or “mistake” in the set-theoretic paradoxes and similar self-referential antinomies is often taken to be the self-reference or self-predication.

In all the above contradictions (which are merely selections from an indefinite number) there is a common characteristic, which we may describe as self-reference or reflexiveness. [36, p. 61]

The iterative notion of a set requires the universal for a property to be of higher type or rank than the instances so that “ $x \in x$ will become a meaningless proposition; and in this way the contradiction is avoided.” To avoid the paradoxes, Whitehead and Russell postulated the *vicious circle principle*: “Whatever involves all of a collection must not be one of the collection.” (Whitehead and Russell [36, p. 37]) But a self-predicative universal of category theory is the “One *among* the many” that transfers the property to all the instances so it is “impredicative” or self-predicative⁷ in violation of the vicious circle principle. Indeed, the universals of category theory are *always* self-predicative via the identity morphisms so the question arises of how category theory avoids similar paradoxes.

All morphisms can be seen as “uniquely factoring through” themselves by the identity morphism (at either the head or tail of the arrow) — so the construction of something like “a universal morphism for all those morphisms that don’t factor through themselves” would always come up empty. Abstractly put, there can be no self-predicative universal for the property of not being self-predicative — since the universal needs to have the property that is “transferred” to the instances by their “participation” in the universal and that particular negative property would always be defeated by the universal’s identity morphism. Thus the problem with

⁷ The connection between the self-predicative or “impredicative” definitions (which caused the problems in naïve set theory) and the self-predicative universals of category theory has not escaped the attention of category theorists. For instance, Michael Makkai notes that the “Peano system” of natural numbers is the self-predicative universal for the property of being a “pre-Peano system”: “we can say that a Peano system is distinguished among pre-Peano systems by the fact that it has exactly one morphism to any pre-Peano system. (An ‘impredicative’ definition if there ever was one!)” (Makkai [20, p. 52]) See Mac Lane and Birkhoff for a full explanation of that “Peano-Lawvere Axiom” [19, p. 67] characterizing the Natural Numbers as the self-predicative universal for counting systems.

the paradoxes was *not* the self-predication *per se* but the *negated* self-predication, and that is defeated in category theory by the universals being *always* self-predicative (by the identity morphisms). The “circle” or self-reference is not the problem if all the circles are required to be “virtuous” so that a “vicious” circle cannot arise.

8. The Third Man Argument

Much of the modern Platonic literature on self-participation and self-predication (Nehamas [25]; Allen [2]; Malcolm [22]) stems from the work of Geach [9] and Vlastos [33] on the Third Man Argument. The name derives from Aristotle (who denied self-predication), but the argument occurs in the dialogues.

But now take largeness itself and the other things which are large. Suppose you look at all these in the same way in your mind’s eye, will not yet another unity make its appearance — a largeness by virtue of which they all appear large?

So it would seem.

If so, a second form of largeness will present itself, over and above largeness itself and the things that share in it, and again, covering all these, yet another, which will make all of them large. So each of your forms will no longer be one, but an indefinite number.

[Parmenides, 132]

If a form is self-predicative, the participation relation can be interpreted as “resemblance.” An instance has the property F because it resembles the paradigmatic example of F -ness. But then, the Third Man Argument contends, the common property shared by Largeness and other large things gives rise to a “One over the many”, a form Largeness* such that Largeness and the large things share the common property by virtue of resembling Largeness*. And the argument repeats itself giving rise to an infinite regress of forms. A key part of the Third Man Argument is what Vlastos calls the Non-Identity thesis:

NI If anything has a given character by participating in a Form, it is not identical with that Form. [33, p. 351]

It implies that Largeness* is not identical with Largeness.

P. T. Geach [9] has developed a self-predicative interpretation of Forms as standards or norms, an idea he attributes to Wittgenstein.

A stick is a meter long because it resembles, lengthwise, the standard meter measure. Geach avoids the Third Man regress with the exceptionalist device of holding the Form “separate” from the many so they could not be grouped together to give rise to a new “One over the many.” Geach aptly notes the analogy with Frege’s *ad hoc* and unsuccessful attempt to avoid the Russell-type paradoxes by allowing a set of all and only the sets which are not members of themselves — except for that set itself (Quine [26]; Geach [10]).

The category-theoretic model for self-predicative universals shows that the flaw in the Third Man argument lies not in self-predication but in the Non-Identity thesis. “The One” is not necessarily “over the many”; it can be (isomorphic to) one among the many. In the special case of sets ordered by inclusion, the union or intersection of a collection of sets is not necessarily distinct from the sets in the collection; the “One” could be one among the many.

For example, let $A = \bigcup\{A_\beta\}$ be the One formed as the union of a collection of many sets $\{A_\beta\}$ [like $\wp(a \cap b)$]. Then add A to the collection and form the new One* as

$$A^* = \bigcup\{A_\beta\} \cup A.$$

This operation leads to no Third Man regress since $A^* = A$.

9. Conclusion

Whitehead described European philosophy as a series of footnotes to Plato, and the Theory of Forms was central to Plato’s thought. We have seen that the self-predicative universals of category theory provide a rigorous mathematical model for the self-predicative strain in Plato’s Theory of Forms and for the intuitive notion of a concrete universal or paradigmatic exemplar elsewhere in philosophy, literature, and ordinary thought.

Moreover, the always-self-predicative universals of category theory “complete the picture” of the antinomies in Frege’s Paradise; one cannot have one theory of universals that could be either non-self-predicative or self-predicative. The “problem” in the antinomies was not the self-predication *per se* but allowing *negated* self-predication. That leaves two ways to avoid the antinomies in a theory of universals, have never-self-predicative universals in the presence of a negation operation (i.e., in set

theory) or always-self-predicative universals without negation (category theory). The always-self-predicative universals of category theory form the complementary bookend to the never-self-predicative universals of set theory — which at least at the conceptual level should foster some peaceful coexistence between these two foundational theories.

References

- [1] Allen, R. E., “Participation and predication in Plato’s Middle Dialogues”, *The Philosophical Review* 69, 2 (1960): 147–164. DOI: [10.2307/2183501](https://doi.org/10.2307/2183501)
- [2] Allen, R. E. (ed.), *Studies in Plato’s Metaphysics*, London: Routledge & Kegan Paul, 1965.
- [3] Awodey, S., *Category Theory* (Oxford Logic Guides), Oxford University Press, Oxford, 2006. DOI: [10.1093/acprof:oso/9780198568612.001.0001](https://doi.org/10.1093/acprof:oso/9780198568612.001.0001)
- [4] Boolos, G., “The iterative conception of set”, *The Journal of Philosophy* 68, 8 (1971): 215–231. DOI: [10.2307/2025204](https://doi.org/10.2307/2025204)
- [5] Desmond, W., *Art and the Absolute: A Study Of Hegel’s Aesthetics*, Albany: State University of New York Press, 1986.
- [6] Ellerman, D., “Category theory and concrete universals”, *Erkenntnis* 28 (1988): 409–429.
- [7] Fujisawa, N., ‘ $E\chi\epsilon\iota\nu$, $M\epsilon\tau\epsilon\chi\epsilon\iota\nu$, and idioms of ‘paradeigmatism’ in Plato’s theory of forms”, *Phronesis* 19, 1 (1974): 30–58.
- [8] Galluzzo, G., and M. J. Loux (eds.) *The Problem of Universals in Contemporary Philosophy*, Cambridge UK: Cambridge University Press, 2015. DOI: [10.1017/CB09781316181539](https://doi.org/10.1017/CB09781316181539)
- [9] Geach, P. T., “The Third Man Again”, *The Philosophical Review* 65, 1 (1956): 72–82. DOI: [10.2307/2182189](https://doi.org/10.2307/2182189)
- [10] Geach, P. T., *Logic Matters*, Berkeley: University of California Press, 1980.
- [11] Hatcher, W., *The Logical Foundations of Mathematics*. Oxford: Pergamon Press, 1982.
- [12] Honderich, T. (ed.) *The Oxford Companion to Philosophy. New Edition*, Oxford UK: Oxford University Press, 2005.
- [13] Hungerford, T. W., *Algebra*, New York: Springer-Verlag, 1974.
- [14] Kneale, W., and M. Kneale, *The Development of Logic*, Oxford: Oxford University Press, 1962.
- [15] Lawvere, F. W., “Adjointness in foundations”, *Dialectica* 23, 3–4 (1969): 281–295. DOI: [10.1111/j.1746-8361.1969.tb01194.x](https://doi.org/10.1111/j.1746-8361.1969.tb01194.x)
- [16] Lawvere, F. W., and S. Schanuel, *Conceptual Mathematics: A First Introduction to Categories*, New York: Cambridge University Press, 1997. DOI: [10.1017/CB09780511804199](https://doi.org/10.1017/CB09780511804199)

- [17] Mac Lane, S., “Groups, categories, and duality”, *Proc. Nat. Acad. Sci. U.S.A.* 34, 6 (1948): 263–267.
- [18] Mac Lane, S., *Categories for the Working Mathematician*. New York: Springer-Verlag, 1971. DOI: [10.1007/978-1-4757-4721-8](https://doi.org/10.1007/978-1-4757-4721-8)
- [19] Mac Lane, S., and G. Birkhoff, *Algebra* (First Edition), New York: Macmillan, 1967.
- [20] Makkai, M., “Structuralism in Mathematics”, pages 43–66 in R. Jackendoff, P. Bloom, and K. Wynn (eds.), *Language, Logic, and Concepts: Essays in Memory of John Macnamara*, Cambridge: MIT Press (A Bradford Book), 1999.
- [21] Magnan, F., and G.E. Reyes, “Category theory as a conceptual tool in the study of cognition”, pages 57–90 in J. Macnamara and G.E. Reyes (eds.), *The Logical Foundations of Cognition*, New York: Oxford University Press, 1994.
- [22] Malcolm, J., *Plato on the Self-Predication of Forms*, Oxford: Clarendon Press, 1991. DOI: [10.1093/acprof:oso/9780198239062.001.0001](https://doi.org/10.1093/acprof:oso/9780198239062.001.0001)
- [23] Marquis, J.-P., “Three kinds of universals in mathematics”, pages 191–212 in B. Brown and J. Woods (eds.), *Logical Consequence: Rival Approaches and New Studies in Exact Philosophy: Logic, Mathematics and Science*, Vol. II, Oxford: Hermes, 2000.
- [24] Miles, M. R., *The Word Made Flesh: A History of Christian Thought*, Malden MA: Blackwell, 2005.
- [25] Nehamas, A., “Self-predication and Plato’s theory of forms”, *American Philosophical Quarterly* 16, 2 (1979): 93–103.
- [26] Quine, W. V. O., “On Frege’s Way Out”, *Mind* 64, 254 (1955): 145–159. DOI: [10.1093/mind/LXIV.254.145](https://doi.org/10.1093/mind/LXIV.254.145)
- [27] Quine, W. V. O., *Mathematical Logic*, Cambridge, MA: Harvard University Press, 1955.
- [28] Russell, B., *Principles of Mathematics*, London: Routledge Classics, 2010 (1903).
- [29] Samuel, P., “On universal mappings and free topological groups”, *Bull. Am. Math. Soc.* 54, 6 (1948): 591–598. DOI: [10.1090/S0002-9904-1948-09052-8](https://doi.org/10.1090/S0002-9904-1948-09052-8)
- [30] Sayre, K., *Metaphysics and Method in Plato’s Statesman*, New York: Cambridge University Press, 2006.
- [31] Shoenfield, J., *Mathematical Logic*, Reading MA: Addison-Wesley, 1967.
- [32] Stern, R., “Hegel, British idealism, and the curious case of the concrete universal”, *British Journal for the History of Philosophy* 15, 1 (2007): 115–153. DOI: [10.1080/09608780601088002](https://doi.org/10.1080/09608780601088002)
- [33] Vlastos, G., . *Platonic Studies*, Princeton NJ: Princeton University Press, 1981.

- [34] Vlastos, G., *Studies in Greek Philosophy. Volume II: Socrates, Plato, and Their Tradition*, edited by D. W. Graham, Princeton NJ: Princeton University Press, 1995.
- [35] Vlastos, G. (ed.) *Plato: A Collection of Critical Essays I: Metaphysics and Epistemology*, Notre Dame: University of Notre Dame Press, 1978. DOI: [10.1007/978-1-349-86203-0](https://doi.org/10.1007/978-1-349-86203-0)
- [36] Whitehead, A. N., and B. Russell, *Principia Mathematica to *56*, Cambridge UK: Cambridge University Press, 1997 (1910).
- [37] Wimsatt Jr., W. K., “The structure of the ‘concrete universal’ in literature”, *PMLA* 62 (1947): 262–280.

DAVID ELLERMAN
Department of Philosophy
University of California at Riverside
4044 Mt. Vernon Ave
Riverside, CA 92507 USA
david@ellerman.org