Benacerraf’s Mathematical Antinomy

Brice Halimi

Paris West University (IREPH) & SPHERE

1 Benacerraf’s dilemma about mathematics as a mathematical antinomy

Benacerraf’s paper entitled “Mathematical Truth” takes on the form of a well-known dilemma. Either a uniform semantics for ordinary language is extended to mathematical language, but then one lapses into platonism; Or a reasonable epistemology of mathematical knowledge as a proof activity is put forward, but then no account of mathematical truth as truth is given. Of course, the semantic horn and the epistemic horn are not exactly on a par, but Benacerraf’s paper precisely tends to tilt the scales so as to get an embarrassing balance (a dilemma). The two horns of the dilemma are mutually exclusive and, to the extent that they are both intenable, Benacerraf leaves it to the reader to understand that they are both wrong.

Such a situation is quite redolent, it will be argued, of the paradigmatic opposition of two equally false theses, that constitutes the core of Kant’s antinomies of pure reason, and more precisely of the first two antinomies, the so-called “mathematical antinomies,” whose both opponents, Kant argues, are wrong —whereas both opponents of the two last antinomies (the “dynamical” ones) are right, albeit from two different points of view. Benacerraf’s dilemma could seem to be a dynamical antinomy, as though mathematical truth could be looked at from the point of view of epistemology as well as from that of semantics, so that both claims would be legitimate within their respective limits. But, against that view, Benacerraf’s dilemma must clearly be understood as an antinomy of the mathematical kind, where both opposite claims are false, as Benacerraf makes it plain they are.

A comparison between Benacerraf’s dilemma and Kant’s mathematical antinomies is called for by strong analogies, as will be seen below, but it has also a purpose. It is indeed a natural question to ask what Benacerraf’s dilemma is driving at, since Benacerraf’s text does not provide any clear solution. On that score, a comparison with Kant’s Transcendental Dialectic could turn out to be
useful, since Kant, in addition to presenting a predicament analogous to Benacerraf’s dilemma, does provide a solution for it. The aim of the present paper is to harness the analogy so as to explore the possibility of transposing Kant’s solution to Benacerraf’s setting.

Let’s first recall the two first antinomies of pure reason. The first one lies in the conflict of the two following theses. Thesis: “The world has a beginning in time, and in space it is also enclosed in boundaries.” Antithesis: “The world has no beginning, and no bounds in space, but is infinite with regard to both time and space.”1 As to the second antinomy, its thesis is: “Every composite substance in the world consists of simple parts, and nothing exists anywhere except the simple or what is composed of simples.” On the contrary, its antithesis claims: “No composite thing in the world consists of simple parts, and nowhere in it does there exist anything simple.”2

In all Kantian antinomies, each claim is mainly negative: It relies on a reductio ad absurdum and feeds entirely upon the impossibility of the opposite claim. In the same way, each horn of Benacerraf’s dilemma draws its strength only from the predicament of the other. So the main structure of the argumentation is the same in both cases. Now let’s delve a bit into the particulars of the antinomy of pure reason. The antinomy, as the whole dialectic of pure reason, is no accident, to begin with. Pure reason is bound to meet contradictions as soon as it is applied to objects of experience. Indeed, a concept of reason (such as the concept of the world) seeks what is unconditioned with respect to some condition and, for that purpose, carries out the synthesis of the regressive series of conditions for a given conditioned, or rather takes that synthesis for granted (already completed) and presupposes the absolute totality of the series of the conditions. So there are in fact two different ways of conceiving of the unconditioned3: either as being the last term of the regressive series of conditions, or as consisting in the whole series itself.

In each antinomy, the thesis epitomizes the first conception, the antithesis the second one. The thesis seeks a first unconditioned entity on which the whole series of conditions depends: It is the reason trying to catch up with the understanding, since the unconditioned is presented as an actual object (although it is prevented from being an empirical one). In a reversal from that, the antithesis presents the absolute sum itself of the conditions in the series as constituting an unconditioned totality: It is the understanding trying to catch up with reason, since it extends the successive synthesis of appearances (empirically discharged by the understanding) into an ideally completed synthesis of the whole series of conditions. As a

1Kant [1998], A426-427/B454-455, p. 470-471.
consequence, cosmological ideas are either too large or too small for the empirical regress sustained by the concepts of the understanding. The first antinomy is an example of this discrepancy:

[Assume] that the world has no beginning: then it is too big for your concept; for this concept, which consists in a successive regress, can never reach the whole eternity that has elapsed. Suppose it has a beginning, then once again it is too small for your concept of understanding in the necessary empirical regress. For since the beginning always presupposes a preceding time, it is still not unconditioned, and the law of the empirical use of the understanding obliges you to ask for a still higher temporal condition, and the world is obviously too small for this law.⁴

Hence each antinomy originates from the mismatch between understanding and reason. In each case, the antithesis sticks to the limits of sensible experience given by the conditions of space and time; On the contrary, the thesis goes beyond those limits and aims at some given absolute entity, whose purported existence is a demand of reason to make sense of a totality of conditions, starting with some given conditioned. Owing to those orientations, Kant associates the thesis of each antinomy to dogmatism, and the antithesis to empiricism.⁵ With the empiricism embodied by the antithesis, “the understanding is at every time on its proper ground, namely the field solely of possible experiences⁶:” The connections between appearances and the laws of those connections are the main focus. On the contrary, the dogmatism embodied by the thesis expresses the voice of reason in its quest for an unconditioned entity in individuo —what Kant calls an Object, as opposed to a Gegenstand.

At this point, an analogy with Benacerraf’s dilemma suggests itself. How far it goes has for sure to be examined, but, at least, the analogy is natural: between the stress put by empiricism on the understanding and the stress put by the epistemic horn on knowledge; between the stress put by dogmatism on Objects and the stress put on platonism on mathematical objects. Of course, both contexts differ substantially. But, precisely, an analogy does not mean a comparison. The whole

⁵See Kant [1998], A465-466/B493-494, p. 498: “In the assertions of the antithesis, one notes a perfect uniformity in their manner of thought and complete unity in their maxims, namely a principle of pure empiricism, not only in the explanation of appearances in the world, but also in the dissolution of the transcendental ideas of the world-whole itself. Against this the assertions of the thesis are grounded not only on empiricism within the series of appearances but also on intellectualistic starting points, and their maxim is to that extent not simple. On the basis of their essential distinguishing mark, however, I will call them the dogmatism of pure reason.”
⁶Kant [1998], A468/B496, p. 499.
working hypothesis of this paper, far from any direct resemblance, is the analogy relating, on the one hand, the function of the understanding and the function of mathematical proofs; on the other, the function of entities of pure reason and the function of mathematical objects. As one shall see, the analogy goes deep and leads one to a solution of Benacerraf’s dilemma.

Kant’s *Critique* characterizes the understanding as the faculty of relations and laws, and reason as the faculty of absolute entities. In the same way, the epistemic horn of Benacerraf’s dilemma belongs in the camp of formal proofs and syntactic rules, whereas the semantical horn epitomizes the quest for an absolute mathematical referent. Empiricism, Kant says, encourages and furthers knowledge, but at the cost of the practical (in the Kantian sense), whereas dogmatism, especially Platonism, meets the practical interest of reason but neglects the investigation of natural appearances. In the same way, the epistemic horn focuses on proofs and on mathematical activity taken in itself, at the cost of mathematical referentiality, whereas the semantic horn meets the practical need of a unified referential framework, but lets the objects prevail over one’s possible access to them.

Thus, the analogy leading from the antinomy of pure reason to Benacerraf’s dilemma consists in connecting the Kantian understanding with the production of deductive series, and the Kantian reason with the reference to absolute entities underlying those series; in relating the platonism of the thesis to Gödelian platonism and the empiricism of the antithesis to Hilbertian finitism. Instead of saying that space has a definite extension, the dogmatic, in Benacerraf, claims that mathematical objects are definite actual denotations. To get back to the first antinomy, instead of saying that space is boundless, the empiricist claims that a mathematical object is nothing but the illimited sum of all the formal proofs that we can produce about its symbol, and does not exist beyond those proofs.

There are in fact various precise similarities between the respective situations described by Kant and Benacerraf, which urge to reconsider “Mathematical truth” as a genuine Dialectic of Mathematical Reason. The table below summarizes the main similarities:

---


8Admittedly, both opponents of Kant’s first antinomy agree to refer to “the world,” whereas the formalist, in Benacerraf’s dilemma, is unwilling to acknowledge any mathematical referent whatsoever. This seems to weaken the comparison between the empiricist antithesis in Kant and the epistemic horn in Benacerraf. In fact, this should rather urge one to understand how inconsistent the concept of world is, as the empiricist describes it, namely as being both an *infinite* series and a *given* totality (see Kant [1998], A418/B445-446, p. 465). Whereas Benacerraf’s formalist claims not to refer to anything, Kant’s empiricist claims to refer to something which actually cannot be anything. This difference qualifies the parallel, but to only a very limited extent, because the empiricist of the first antinomy does not invoke any individual entity —only an infinite series which, even if it is supposed to be completed, can hardly be compared to a referent in the usual sense of the word.
<table>
<thead>
<tr>
<th><strong>Kant</strong></th>
<th><strong>Benacerraf</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>mathematical antinomy</td>
<td>dilemma</td>
</tr>
<tr>
<td>antithetic (each thesis feeds upon the contradiction of the other)</td>
<td>negative argumentation</td>
</tr>
<tr>
<td>understanding</td>
<td>“our ability to produce and survey formal proofs” (p. 668)</td>
</tr>
<tr>
<td>appearances</td>
<td>symbols</td>
</tr>
<tr>
<td>possible experience</td>
<td>admissible inference</td>
</tr>
<tr>
<td>series of conditions</td>
<td>deductive chains</td>
</tr>
<tr>
<td>law</td>
<td>theorem</td>
</tr>
<tr>
<td>reason</td>
<td>truth theory</td>
</tr>
<tr>
<td>unconditioned being</td>
<td>direct reference</td>
</tr>
<tr>
<td>intellectual intuition</td>
<td>Gödelian intuition</td>
</tr>
<tr>
<td>dogmatism (thesis)</td>
<td>realism (standard conception)</td>
</tr>
<tr>
<td>Platonism</td>
<td>platonism (in mathematics)</td>
</tr>
<tr>
<td>empiricism (antithesis)</td>
<td>finitism (combinatorial conception)</td>
</tr>
<tr>
<td>Epicureanism</td>
<td>Hilbertian formalism</td>
</tr>
<tr>
<td>The Idea is either too big or too small for the concept of the understanding (A486/B514, p. 508).</td>
<td>“[…] squeezing the balloon at that point [that of knowledge] apparently makes it bulge on the side of truth.” (p. 668)</td>
</tr>
</tbody>
</table>
| The thesis favors the practical and is more popular for that reason (A466-467/B494-495, p. 498-499). | The semantic horn accounts for mathematical truth and dovetails with the pragmatic needs of ordinary communication. | 9
| solution provided by transcendental idealism   | to be specified                                      |

9 In Benacerraf [1965] (p. 71-72), Benacerraf considers shifting back from “the” natural numbers to the space of all \( \omega \)-progressions, but remarks that “ordinary communication” requires one single distinguished sequence of notations. This is echoed by the semantic horn in Benacerraf [1973]: For practical reasons, we need to speak of “the” natural numbers. Yet we should step back from that habit if we wish to be faithful to the way mathematics really works: Speaking of the natural numbers is really just a way of speaking of any \( \omega \)-progression. The structure of
Further study of Kant’s Transcendental Dialectic has to be foregone for the moment, but the detour through Kant calls forth one important point: It hints at a solution of Benacerraf’s dilemma, since Kant provided a systematic and extensive solution for the antinomies. In fact, Kant’s solution is in some respects redesigned by Benacerraf himself, although not in Benacerraf [1973]. This is what can be established on the basis of two other papers from Benacerraf, Benacerraf [1965] and Benacerraf [1981]. Thus, before turning toward the possible transposition of Kant’s solution to Benacerraf’s problem, those two papers must be examined. They will confirm that Benacerraf’s dilemma is akin to a mathematical antinomy, as opposed to a dynamical one, i.e., that the two views in conflict cannot be conciliated but must be both overcome. Then one will consider the transposition of Kant’s solution, and to what extent Benacerraf carried it himself out.

2 Bring the kids and the founding father on

This section aims to show, first, that Benacerraf [1965] proves the semantic horn to be inconsistent and then, going forward up, that Benacerraf [1981] proves the epistemic horn to be inconsistent.

2.1 Backward to Benacerraf [1965]: Objects involve proofs

If Benacerraf [1973] foregrounds the semantic constraint, Benacerraf [1965] clearly shows that the semantic values of numerical terms like ‘3’ or ‘17’ are not univocal. Even when we are using genuine singular terms in mathematics, their reference is not set unambiguously. So the basic lesson to be learned is that neither Ernie’s account nor Johnny’s account can provide the right semantic value for natural numbers. On that score, neither Ernie nor Johnny succeeds in securing any of the two horns.

In fact, Ernie’s account (i.e., Zermelo’s) and Johnny’s (i.e., von Neumann’s) can be construed as being, precisely, two different interpretations of the same mathematical objects. A mathematical object like 3 is in fact a series of objects (containing, to begin with, {∅}, {{∅}}, {∅, {{∅}}}) as well as {{∅}}), supplemented with the proof that the models to which they belong respectively are mutually interpretable.

Let’s consider $E$ and $J$, respectively the “Ernie interpretation” and the “Johnny interpretation” of $L = \{\in, 0, S\}$: $E$ is the L-structure whose domain is natural numbers is an abstraction that results from the mere desire to single out one privileged $\omega$-progression. From that point of view, as Benacerraf [1965] points out, the main misconception about mathematical objects is that we conceive of each of them as being, at the same time, a proxy for a whole class of particular objects, and a particular member of that class as well.
\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \ldots\} and J is the L-structure whose domain is \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \ldots\}. An interpretation of J in E consists:

- in a formula \(\partial(x)\) of L
- for each atomic formula \(\phi(y_0, \ldots, y_{m-1})\) of L, in a formula \(\phi^* (x_0, \ldots, x_{m-1})\) of L
- in a surjective map \(f : \partial^E \subseteq |E| \rightarrow |J|\) such that, for all atomic formula \(\phi(x_0, \ldots, x_{m-1})\) of L and all \(a_i \in \partial^E\) (\(0 \leq i \leq m - 1\)):
  \[E \models \phi^* [a_0, \ldots, a_{m-1}] \text{ iff } J \models \phi [f(a_0), \ldots, f(a_{m-1})].\]

The formula \(\partial\) is called the \textit{domain formula} of the interpretation, and the map \(f\) is called its \textit{coordinate map}: It assigns to each element \(f(a)\) in \(|J|\) its “coordinate” \(a\) in \(|E|\).

There is a straightforward interpretation of J in E, given by:

- \(\partial(x)\) is \('x = x'\), so that \(\partial^E = |E|\);
- \(f(0^E) = 0^J\) and \(f(n \cup \{n\}) = \{n\}\), so that \(f(S(n)^E) = S(n)^J\);
- \((x \in y)^* = x \in^* y\), where \(\in^*\) is the ancestral relation of \(\in\). This amounts to \(x \in TC(y)\), where \(TC(y)\) is the transitive closure of \(y\).

Let ZFC\(^+\) be the theory ZFC extended by the definition of the function symbol \('TC(y)'\); ZFC\(^+\) is a conservative extension of ZFC — any model of ZFC can be expanded to a model of ZFC\(^+\).

More generally, \(\phi^* = \phi[\in^* / \in]\).

One can check that \(E \models x \in y [n, p]\) iff \(J \models x \in^* y [n, p]\). For instance, \(E \models 0 \in 2\), and \(J \models 0 \in^* 2\) because \(J \models (0 \in 1 \land 1 \in 2)\).

So natural numbers correspond in fact to a series of models, \textit{supplemented with the proof} that any two models of the series are mutually interpretable. This is the way to exhibit “\textit{the} natural numbers” as the invariant of the series. If we have canonical interpretations (canonical coordinates) between any two models, then we can speak, in the same way, of the series of all realizations of “\textit{the number 3}.” Quite generally, a mathematical object is but the invariant of the series of all its possible realizations or “presentations,”\(^{10}\) which involves in a crucial way the \textit{proof} of their mutual interpretability into one another. The proof is built into the mathematical object \textit{qua} mathematical. So the semantic horn is actually inconsistent.

\(^{10}\)About the notion of “presentation,” see below, p. 12.
Further misgivings about the semantic horn should be mentioned. First of all, explaining the referentiality of a mathematical theory, namely granting that that theory is not simply wheelspinning, does not require to abide by the superficial grammar. And Benacerraf knows that very well. But we can leave that issue aside, to the extent that the problem of the semantical value remains in the end.

A more serious concern is the following. The semantical constraint falls to the philosopher of mathematics. It behooves her as a philosopher specifically, to explain the semantical value of mathematical terms, so as to account for the truth of mathematical sentences molded upon some kind of correspondence (in the framework of Tarski’s convention T). Now, the problem is very simple: Tarskian semantics, which is explicitly taken as the scheme to comply with, relies heavily on a background set-theoretic universe that is taken for granted. The members of that universe are mathematical objects. In other words, Tarskian semantics is mathematical in nature.

Admittedly, as to the sentence “Socrates is Plato’s master,” ‘Socrates’ is interpreted by the individual Socrates himself, not by any mathematical entity whatsoever. But what about “All men are mortal?” In Tarskian semantics, the interpretation of such a sentence calls at least for the domain of all men, which is a set— with all the human beings as urelemente. So it is not true that, in the perspective of the semantic horn, mathematical objects come into play for the interpretation of mathematical terms only: They come into play for the interpretation of about any sentence. Mathematical objects are resorted to, not only to account for mathematical truth, but, more generally, to account for truth simpliciter.

But now, the actual semantical values of the set-theoretic terms that any Tarskian-style semantics resorts to, should be accounted for themselves. The mathematical terms of the semantical metalanguage ought to be properly interpreted. So the problem is only pushed back one square higher up and an infinite regress is clearly looming up here. And since Benacerraf’s epistemic point is about any kind of mathematical object, this problem applies not only to set theory strictly construed, but to all branches of mathematics, and to arithmetic, to begin with. Worse, that infinite regress is of the vicious kind, to refer to Russell’s distinction: “[...] whenever the meaning of a proposition is in question, an infinite regress is objectionable, since we never reach a proposition which has a definite meaning.”

2.2 Forward to Benacerraf [1981]: Proofs involve objects

Let’s now turn to Benacerraf’s later argument to the effect that the epistemic horn is inconsistent. In Grundlagen (§3), Frege claims that “the question [as to whether 

12Russell [1903], §329, p. 349.
a proposition is *a priori* or not] is removed from the sphere of psychology, and assigned, if the truth concerned is a mathematical one, to the sphere of mathematics.” Here is Benacerraf’s comment:

Since arithmetical propositions are at issue, the question of their justification is properly a matter for mathematics. Therefore, the concepts will be so defined as to make it a properly *mathematical* question whether some arithmetical judgment is analytic or synthetic, *a priori* or *a posteriori*. [...]

To determine whether a proposition is analytic, look for a proof of it in which the basic propositions are “primitive truths” — propositions which themselves have no proofs. If there exists such a proof (one in which appeal is made only to definitions and to “primitive truths”) and the primitive truths evoked include only laws of logic, the proposition in question is analytic. If not, it is synthetic.13

Benacerraf extends Frege’s claim with the conclusion that the whole enterprise of the *Grundlagen* is “first and foremost a mathematical one.”14 Along with that assessment of Frege’s perspective, it is also important to mention, about the §3 of the *Grundlagen*, that the question as to whether a given mathematical proposition is analytic becomes a mathematical one, not only because “the truth concerned is a mathematical one,” but because the way of establishing its analyticity is mathematical. The construction and review of all the steps of a mathematical truth is a mathematical task —ideography being exactly the tool geared to that task. So only a mathematician *qua* mathematician is able to deem a given proposition analytical or synthetical. The rigorous representation of a mathematical proof is a piece of mathematics in itself, that is the implicit part of Frege’s claim that “the question is assigned to the sphere of mathematics.”

Benacerraf insists on Frege’s theory being a theory of analytical truth. The further point that should be made here is that, precisely, Frege’s theory is a mathematical theory, laid out in the ideographical language. The recognition of special mathematical truths as being analytical requires to turn proofs themselves into mathematical objects in their own right. This is also the reason why the hierarchy of truths is, in Frege’s reckoning, completely objective. So the conclusion, following Frege’s perspective, is that mathematical proofs, even viewed as purely formal, have to be recognized as mathematical objects in their own right, which are no less epistemically problematic than the number 3. To that extent, the epistemic horn is actually inconsistent.

14Benacerraf [1981], p. 34.
In his discussion of the status of Frege’s definitions, Benacerraf [1981] adds a further point. About the fact that, in Frege’s *Grundgesetze*, any course-of-values can be taken to be the True, Benacerraf writes:

> Of course it does not make any mathematical difference. But *that* it makes no mathematical difference is an important philosophical point concerning what we must construe definitions such as Frege’s to accomplish. Although I cannot pursue the matter further here, I hope that these examples make it clear that a straightforwardly “realist” construal of Frege’s intentions or accomplishments will fail to do justice to his practice.  

If one gathers this argument with the discussion of the accounts of numbers by Ernie and Johnny, two kinds of situations emerge where one has to deal with two mathematically equivalent items: either, as in the case of Ernie and Johnny, two different tokens of the same mathematical character, to use Kaplan’s terminology; or, as in the case of the referent of the True, two different choices that amount to the same thing mathematically speaking. In both cases, a mathematical referent emerges as a cluster of proofs —the proofs that one is dealing with different but mathematically equivalent presentations of what has to be recognized then as a mathematical invariant.

In his 1965 paper, Benacerraf does not intend to build up a predicament, but precisely shows instead, in a positive way, that a mathematical object cannot but be the invariant of its different presentations, and as such involves the proof that any two such presentations are equivalent (in the relevant sense of “equivalent”); That equivalence proof is built into the object properly understood as a mathematical object. In his 1981 paper, Benacerraf stresses that a mathematical definition need preserve neither the meaning nor even the reference of the *definiendum*, because it is meant to introduce a mathematical object whose identity conditions are determined by what one needs to prove with or about it. In both cases, the dichotomy between mathematical objects and formal proofs, that is instrumental in setting up Benacerraf’s dilemma, has to be overcome. *Above all*, the fact that a mathematical object consists properly of the system of all its different presentations (this word will soon be got back to) and generally cannot be mentioned apart from one of these, is fully recognized by Benacerraf, and acknowledged by him in a much more positive way than it is usually made out to be.

---

Benacerraf [1981], p. 31.
3 Back to Kant

3.1 The analogy put to work

One of the main benefits that can be expected from the analogy made between Kant’s *Dialectic* and Benacerraf’s dilemma is to provide a clue for a solution of the latter. What actually is Kant’s solution of the whole antinomy of pure reason? Let’s consider again the case of the first antinomy:

If one regards the two propositions ‘The world is infinite in magnitude,’ ‘The world is finite in magnitude,’ as contradictory opposites, then one assumes that the world (the whole series of appearances) is a thing in itself. […] But if I take away this presupposition, or rather this transcendental illusion, and deny that it is a thing in itself, then the contradictory conflict of the two assertions is transformed into a merely dialectical conflict, and because the world does not exist at all (independently of the regressive series of my representations), it exists neither as *an in itself infinite* whole nor as *an in itself finite* whole. It is only in the empirical regress of the series of appearances, and by itself it is not to be met with at all.

[…]

Accordingly, the antinomy of pure reason in its cosmological ideals is removed by showing that it is merely dialectical and a conflict due to an illusion arising from the fact that one has applied the idea of absolute totality, which is valid only as a condition of things in themselves, to appearances that exist only in representation, and that, if they constitute a series, exist in the successive regress but otherwise do not exist at all.16

As a consequence, the series of conditions for a given conditioned should be understood, not as a regress in infinitum (regress to infinity), but as a regress in indefinitum (indeterminately continued regress) whose absolute completion cannot be postulated:

[The regress in the series of conditions] is a principle of the greatest possible continuation and extension of experience, in accordance with which no empirical boundary would hold as an absolute boundary; thus it is a principle of reason which, as a rule, postulates what should be effected by us in the regress, but does not anticipate what is given in itself in the object prior to any regress.17

17Kant [1998], A509/B537, p. 520. See also A518-523/B546-551, p. 525-528.
The whole antinomy, and the entire Dialectic as well, in fact, relies on the false assumption that the objects to which refer the ideas of reason are given in themselves, whereas they are given only in the course of a regressive series of conditions:

[... ] if it is said that the world is either infinite or finite (not-infinite), then both propositions could be false. For then I regard the world as determined in itself regarding its magnitude, since in the opposition I not only rule out its infinitude, and with it, the whole separate existence of the world, but I also add a determination of the world, as a thing active in itself which might likewise be false, if namely, the world were not given at all as a thing in itself; and hence, as regards its magnitude, neither as infinite nor as finite.\(^{18}\)

Now, if we are willing to pursue the analogy established in the beginning of this paper, the dilemma urges to jettison the presupposition that mathematical objects are things in themselves, and that the question of their status calls for a clear-cut answer, namely “they do exist in the sense of a full-fledged existence” or “they are mere ideal fictions in the course of a proof and do not exist at all.” Let’s try to implement that idea in a more precise way. How to make sense of the claim that mathematical objects are not given in themselves? What does it mean? It means something quite basic, namely that any mathematical object goes with presentations whose nature depends on the kind of object at stake. Various examples of multiple presentations abound in mathematics: A vector space is usually presented through an affine space fixed by the arbitrary choice of some origin; The symmetric group \( S_n \) is often described as the group of permutations on \( \{1, 2, \ldots, n\} \), but can equally well be defined as the group of permutations on any other \( n \)-element set; The set \( \mathbb{C} \) of complex numbers can be defined algebraically as \( \mathbb{R}/(X^2 + 1) \), arithmetically as the set \( \{a + ib : a, b \in \mathbb{R}\} \) endowed with addition and multiplication, geometrically as points of the plane; The natural numbers can be defined as Ernie or as John does. Another example is Cantor’s account of ordinals as equivalence classes of well-orderings: It constitutes a certain presentation of the concept of ordinal which itself involves, for each ordinal, a presentation in the form of a representative of this ordinal as equivalence class.

Admittedly, different scales ought to be distinguished, because in some cases the “same” structure is introduced with the help of different supports (as in the case of complex numbers), whereas in some other cases different structures provide different accounts of the “same” mathematical concept (as in the case of the Ernie vs Johnny controversy). All those cases belong to such obviously different scales that the notion of presentation is affected by some irreducible vagueness.

\(^{18}\)Kant [1998], A504/B532, p. 517-518.
which verges on equivocity. For instance, the different presentations of $\mathbb{S}_n$ that have been mentioned are notational variants, whereas the two different presentations of $\mathbb{C}$ are based on altogether different frameworks. This does not detract, however, from the occurrence of the same phenomenon on each scale, namely the diffraction of a “same” mathematical item into various incarnations, without which this mathematical item cannot be grasped, let alone studied. Such is already, in fact, Benacerraf’s diagnosis in Benacerraf [1965]:

> Any purpose we may have in giving an account of the notion of number and of the individual numbers, other than the question-begging one of proving of the right set of sets that it is the set of numbers, will be equally well (or badly) served by any one of the infinitely many accounts satisfying the conditions we set out so tediously.\(^{19}\)

A mathematical object (e.g., a vector space) cannot in general be given “in itself,” canonically, but requires some fixed configuration (e.g., an affine space having that vector space as direction) whose bias enables one to reach the intended structure. Bringing up the notion of mathematical presentation is a general way to single out the pervasive use of such configurations throughout mathematics. Contrary to the actual drawing compared to the geometrical theorem, presentations do not boil down to sub-mathematical conditions of concrete mathematical activity. They are truly mathematical in nature and lend themselves sometimes to an explicit mathematical treatment: The notion of presentation of a group by generators and relations or the notion of resolution of a module are preeminent examples in the field of algebra.

Hence, if we try to make sense of Kant’s solution and to transpose it to Benacerraf’s dilemma, we are lead to the view that a mathematical object is but the invariant of the open-ended series of all its possible presentations, which themselves are neither purely formal items nor independent semantic units. As pointed out above (p. 10), Benacerraf knows that very well, as early as 1965:

> Number theory is the elaboration of the properties of all structures of the order type of numbers. The number words do not have single referents. […] Only when we are considering a particular sequence as being, not the numbers, but of the structure of the numbers does the question of which element is, or rather corresponds to, 3 begin to make any sense.

Slogans like “Arithmetic is about numbers,” “Number words refer to numbers,” when properly urged, may be interpreted as pointing out two distinct things: (1) that number words are not names of special

\(^{19}\)Benacerraf [1965], p. 62.
non-numerical entities, like sets, tomatoes, or Gila monsters; and (2) that a purely formalistic view that fails to assign any meaning whatsoever to the statements of number theory is also wrong.20

Just as the cosmological idea, in Kant, “is only in the empirical regress of the series of appearances, and by itself it is not to be met with at all,”21 in the same way a mathematical structure never exists outside the series of its presentations, none of which can be privileged as giving what would seem to be “the structure itself.” And just as the unreachable completion of a series of conditions is the task prescribed to the understanding as a “regulative principle of reason,” in the same way the never-ending exploration of all the possible presentations of a “same” mathematical structure (definitions becoming theorems, and vice versa), which amounts to the establishment of all the possible theorems about it, is the task, never amenable to completion, that defines mathematics as a discipline.

Just as “the world” really is something, but only as an incomplete series of conditions and as the regulative focus thereof, in the same way a mathematical object actually covers an open-ended bundle of provably equivalent presentations (i.e., an open-ended bundle of equivalence proofs between different presentations), being “itself” only the regulative invariant of this series. It is nothing like a thing-in-itself from which each presentation would originate but, on the contrary, the focal point that multiple presentations project collectively as their common target.

The presentational account defended in this paper claims that mathematical objectivity is intentional, insofar as a coherent bundle of presentations points to an object, but only in a regulative way, without positing any pointee. On the contrary, the semantical account considers that any pointing at all presupposes a pointee that exists by itself, whereas the combinatorial account denies any pointing at all and considers a series of presentations, in and of itself, to be all that there is. But both accounts miss the crucial fact that doing mathematics is constantly shifting from one presentation to another, provably equivalent, one; This fact is not only witnessed by history of mathematics and by mathematical practice, but constitutes the very core and texture of mathematics.

3.2 Presentations

A few words are in order about the notion of presentation as it occurs in the phrase “mode of presentation.” Actually, this phrase can be traced back to Frege. Indeed

---

20Benacerraf [1965], p. 70-71.
21Kant [1998], A505/B533, p. 518. Kant adds: “The series of appearances is to be encountered only in the regressive synthesis itself, but is not encountered in itself in appearance, as a thing on its own given prior to every regress.”
this term was mentioned by Frege in the context of the well-known distinction that he drew between sense and denotation:

If the sign ‘a’ is distinguished from the sign ‘b’ only as object (here, by means of its shape), not as sign (i.e. not by the manner in which it designates something), the cognitive value of \( a = a \) becomes essentially equal to that of \( a = b \), provided \( a = b \) is true. A difference can arise only if the difference between the signs corresponds to a difference in the mode of presentation of that which is designated. Let \( a, b, c \) be the lines connecting the vertices of a triangle with the midpoints of the opposite sides. The point of intersection of \( a \) and \( b \) is then the same as the point of intersection of \( b \) and \( c \). So we have different designations for the same point, and these names (‘point of intersection of \( a \) and \( b \), ‘point of intersection of \( b \) and \( c \)’) likewise indicate the mode of presentation [Art des Gegebenseins]; and hence the statement contains actual knowledge.

It is natural, now, to think of there being connected with a sign (name, combination of words, letter), besides that to which the sign refers, which may be called the reference of the sign, also what I should like to call the sense of the sign, wherein the mode of presentation is contained.\(^{22}\)

The choice that Frege made of the term “presentation” is certainly not insignificant and could certainly be driven back to two opposite sources contemporary with Frege’s work: the work of Franz Brentano in psychology, especially his 1874 book, *Psychology from an empirical Standpoint*\(^ {23} \), where the notion of “mode of presentation” (Modus des Vorstellens) is ubiquitous, in particular in his account of time-consciousness; and the foundation of the theory of group presentations by Walther Dyck (a student of Felix Klein) in his article “Gruppentheoretische Studien,” published in 1882.\(^ {24} \) Admittedly, the terms used are different: Brentano uses “Vorstellen” (or Vorstellung) and Dyck uses “Präsentation,” whereas Frege uses “Gegebensein.” Notwithstanding these terminological discrepancies, the affinity between the concepts put forward is strong enough to permit a comparison. In that perspective, the Fregean notion of mode of presentation should be construed as both epistemological and logical, both as an epistemic access and as a presentation in the mathematical sense of the presentation of a group. The mathematical example taken by Frege in the quote must be understood as a way to introduce a notion that is a common platform to account for

\(^{22}\)Frege [1960], p. 57.
\(^{23}\)Brentano [1973].
\(^{24}\)Dyck [1882].
both the basic phenomenon of meaning in ordinary language and the customary
use of various descriptions of mathematical objects. But this is matter for another
paper.\footnote{25}

The main point is that, as early as 1965, before the dilemma was properly
coined, Benacerraf had already a virtual solution to his 1973 dilemma, even though
he did not bring it out — and that his solution is actually directly analogous to that
proposed by Kant, following the general analogy described at the beginning of
this paper. Getting back to “Mathematical truth,” one can express the mistake
of the epistemic horn as the wrong thesis that mathematical presentations do not
present anything, and the mistake of the semantic horn — the mistake of contem-
porary structuralism\footnote{26} — as the wrong thesis that presentations are mere artefacts
as opposed to the mathematical structures “themselves.”

Of course, a lot remains to be explained about the workings of mathematical
presentations, and about their epistemical accessibility. Such a study cannot be
completed within the limits of this paper. For sure, mathematical presentations
take on various aspects, from the choice of letters for the representation of permu-
tations or Gödel numberings, to group presentations or module resolutions. Their
spectrum is hardly amenable to a single kind and more importantly, as underlined
earlier (p. 12), their different cases belong to heterogeneous scales. Moreover,
presentations are certainly many-layered: The presentation of some mathematical

\footnote{25}Jason Stanley and Timothy Williamson (Stanley and Williamson [2001], p. 427), referring
to the Fregean notion of mode of presentation, have transposed it into a practical context: To
follow their example, the ascription to Hannah of a certain knowing how (knowing how to ride
a bicycle) is described as a “practical mode of presentation” of Hannah’s propositional knowledge
that a particular way of riding a bicycle is a way for her to ride a bicycle (see p. 428–429).
Whereas a linguistic or semantical mode of presentation designates the particular way under which
a proposition is entertained, a practical mode of presentation is a particular way of expressing or
understanding “ways of engaging in actions” (p. 436). The analysis of the notion of practical mode
of presentation is still to be developed, as conceded by the authors (p. 429). This paper claims that
the notion of mathematical presentation is certainly clearer to analyze than that of practical mode
of presentation, in particular because the former does not rely on a mere parallel with the Fregean
one, as the latter does, but corresponds to one of the main cases that Frege had in mind.

\footnote{26}Taking presentations seriously, as essential devices in-between mathematical structures
and the empirical systems that instantiate the latter, would be a way to overcome the “identity prob-
lem” that has been raised against \textit{ante rem} structuralism (see, in particular, Keränen [2006] and
Shapiro [2006]). The identity problem comes from the existence of non-trivial automorphisms in
certain structures (such as conjugacy in the field of complex numbers). The solution that could be
given to that problem is that a non-trivial automorphism of a given structure is always, in fact, an
isomorphism between two presentations associated to that structure (for instance, conjugacy is an
isomorphism between \(\langle \mathbb{C}, i, -i \rangle\) and \(\langle \mathbb{C}, -i, i \rangle\) as presentations of the same structure, which does
not admit itself of any non-trivial automorphism). This solution requires to admit that a structure
is never accessible outside some of its presentations, which clearly qualifies realist structuralism,
still without giving in to the empiricist thesis that one can have access only to systems (concrete
instances of structures).
object can itself be turned into an object w.r.t. some of its own presentations.

Despite this variety, a mathematical presentation always relies on a fixed configuration, laid out in an environment that shows how shifting to another configuration would be possible but incidental to the intended mathematical structure. The mediation thus provided by a presentation shows how to loosen Benacerraf’s dilemma. Group presentations are a good example to consider: \([a, a^n = 1]\) is a presentation of the cyclic group \(\mathbb{Z}/n\mathbb{Z}\) among other possible presentations, and yet produces this cyclic group itself. The presentation is neither a mere sequence of symbols (as the formalist would have it), because it actually presents something; nor a proper name (as the platonist would have it), because it has an internal structure that is by itself informative and does not point to any external object. Above all, tracing back “the” cyclic group with \(n\) elements to the presentation \([a, a^n = 1]\) shows how one can deal (and often deals) with the former on the much more cognitively tractable basis of the latter. Of course, other presentations of \(\mathbb{Z}/n\mathbb{Z}\) are available, in particular presentations that are not presentations by generators and relations; One of the tasks of mathematics is precisely to link all the known presentations together as equivalent presentations of the same “thing.”

The “same” mathematical object entertained by multiple presentations is the same object only derivatively, only because of the provable equivalence of all those presentations. This kind of identity is a source of increased complexity. Indeed, the criteria that rule the equivalence of two different presentations, as well as the logical resources (in the broad sense) available to ascertain it, have varied significantly in history. In particular, too weak resources imply that no connection can ever be made. The equivalence relating certain geometric entities with algebraic equations was beyond consideration before Descartes’ analytic geometry. The proof of the equivalence of mathematical objects in the sense of the existence of an equivalence of categories (such as that existing between the category of Boolean algebras and the category of Stone spaces) required first the development of category theory. The variety of tools to set an equivalence is not only historical, that is, diachronical: For instance, the equivalence of categories is a global way of conceiving of equivalence that differs in that respect from the set theoretic notion of isomorphism, yet both notions belong to contemporary mathematics and cannot be assigned to different periods. Thus, the framework within which the comparison of different presentations can be established in a more or less fine-grained way, and the whole apparatus that Bourbaki described in general terms as “transport of structures,” are themselves part of the process from which stable mathematical object emerge. In other words, stabilized ways of assessing the equivalence of (thereby) stable mathematical objects run concurrently with the emergence of those objects. Of course,\(^{27}\) it always remains an option to maintain

\(^{27}\)This objection was raised by Marco Panza and Achille Varzi in two different ways.
that stable objects come first and that equivalence proofs cannot be integral to them, either because a bundle of proofs never can make up an object or because equivalence proofs presuppose as objects the items of which they show the equivalence. But what is the way out of the difficulties to which platonism leads, then? A more constructive answer to meet the objection is this: Any equivalence proof bears on two mathematical constructions which have already, in a sense, the status of pre-objects, but a genuine mathematical object stands out only after several equivalence proofs have secured some kind of invariant: The mathematical way of looking at a triangle, to take an elementary example, begins with disregarding its size or actual position.

There is one last point that it is useful to clear up. As already mentioned, Kant explains that, in the first antinomy, the idea of the world is too small for the concept of the understanding in the thesis, and too big for it in the antithesis. In the case of Benacerraf’s dilemma, one could be tempted to say the other way around: that in the equivalent of the thesis (the standard conception) the idea of mathematical objectivity is too big for the understanding, and that in the equivalent of the antithesis (the combinatorial conception) it becomes too small. As Benacerraf says:

[...] a typical ‘standard’ account (at least in the case of number theory or set theory) will depict truth conditions in terms of conditions on objects whose nature, as normally conceived, places them beyond the reach of the better understood means of human cognition (e.g., sense perception and the like).

[...] postulational stipulation makes no connection between the propositions and their subject matter — stipulation does not provide truth. At best, it limits the class of truth definitions (interpretations) consistent with the stipulations. But that is not enough.28

How to explain this apparent inversion? In fact, there is none. The idea of mathematical objectivity, as supplied by the standard conception, is really too small, and the same idea, as supplied by the combinatorial conception, really too big. The yardstick, indeed, is not so much our cognitive power as an open-ended series of presentations that mathematicians come up with in history. The standard conception closes the process too early and thus supports too narrow an idea of a mathematical object: The content of a mathematical concept is fixed once for all (the number 3 is ...) although it continues being enriched by new theorems, so that the problem becomes to decide whether one keeps the same object when some really new comprehension of it is put forward. (Think of the parabola when

---

it has been rethought as a scheme, or simply of the number 3 when it has been rethought as an ordinal set.) On the contrary, the combinatorial conception contends that a mathematical object is already nothing else than the (thereby too big) complete series of all that is and will be proved about it, which now raises the symmetric problem of accounting for genuine ruptures in history of mathematics. Both opponents neglect the deep historical nature of mathematical objectivity—which should come as no surprise, since the main representants of both sides (Tarski and Hilbert, respectively) foregrounded a mainly logical and an-historical view of mathematics.

**Conclusion**

As a starting point, an analogy can be drawn between Kant’s Antinomies and Benacerraf’s dilemma: Dogmatism becomes the semantic horn, empiricism becomes the epistemic horn. Benacerraf’s dilemma can then be reconsidered as a mathematical antinomy about mathematical objectivity. The analogy turned out to be remarkably steady and sharp.

In that perspective, “Mathematical truth” can be read as an attempt at reconstructing a naive philosophy of mathematics in order to better overcome it, and to overcome it quite in the way Kant undertook to overcome rational cosmology. Two other important papers of Benacerraf’s, Benacerraf [1965] and Benacerraf [1981], confirm that both horns are inconsistent, and that the antinomy is really mathematical, as opposed to dynamical, to refer to Kant’s terminology.

But the main motivation of the analogy remains the prospect of transposing Kant’s solution so as to suggest a way of solving Benacerraf’s dilemma itself. The upshot of the analogy is that a mathematical object can never be given in itself because it consists of the open-ended series of all its possible presentations, in the sense specified above. As an example given by Benacerraf himself, Ernie’s and Johnny’s “accounts” constitute two different presentations of the natural numbers. The standard conception and the combinatorial view are both wrong because they crystallise an open-ended series into either a closed referent or a complete infinite series. They precipitate the dilemma and face further problems, in particular to account for basic historical examples of mutation in the definition of classical mathematical objects. On the contrary, the solution of Benacerraf’s dilemma drawn from the analogy with Kant’s antinomies calls for a more history-sensitive analysis of mathematical objectivity. It acknowledges that mathematical objects do exist, but proposes to conceive of each mathematical object as a series of equivalent presentations which is always in the making, thus revisable, and which cannot be separated from the proofs explaining that the equivalent presentations that it gathers are indeed equivalent, and how they are.
References


