BENACERRAF’S DILEMMA

W. D. HART
University College
London University

Pure mathematics has attracted philosophical attention for as long as philosophy has been written. All the great philosophers seem to have had something to say about mathematics, and some of them, a great deal. Why? It is not, or not just, because of the age and centrality of mathematics in human culture, for agriculture is at least as old and central in civilization, and yet there is next to no philosophy written about it. The reason seems rather to be that there is a problem about mathematics that has engaged philosophers for ages, while no such difficulty peculiar to agriculture seems to arise. Paul Benacerraf made a compelling statement of that problem in 1973,¹ and here we will restate that problem, though not in exactly Benacerraf’s terms. While Benacerraf’s may not be the only philosophical problem about mathematics, it is certainly central and perennial.

Benacerraf’s problem takes the form of a dilemma, that is, of two lines of thought each separately appealing and yet jointly inconsistent. Briefly put, the dilemma is that what is necessary for mathematical truth makes mathematical knowledge impossible. This seems incredible, since from the time of the Greeks, mathematics has been taken as the most absolute body of truth

of which we have the most certain knowledge. Since it is typi-
cally metaphysics that seeks to meet the demands of truth while
it is epistemology that seeks the demands of knowledge, we may
think of Benacerraf's dilemma as having two horns, one meta-
physical, and the other, epistemological.

We begin with the metaphysical horn, which we organize as
an argument for the metaphysical thesis of platonism, that is,
that there are very abstract, non-mental and non-physical, ob-
jects such as the numbers, functions and sets of mathematics.
Our first premiss is a version of the common sense that truth
is correspondence to fact. We take it to be sentences that are
true or false. Perhaps we could think here instead of proposi-
tions, thoughts or statements. But whether it is true that Zapata
wore a mustache seems clearly to depend on Zapata. In gen-
eral, whether a truth vehicle\textsuperscript{2} is true or false seems to depend
on an articulation of that vehicle into units smaller than whole
truth vehicles. Sentences wear an articulation in words on their
inscribed faces. But, unless one just reads the articulation of
a sentence used to express a proposition, thought or statement
back into that proposition, thought or statement, it is not patent
that, or if so, how, such things are segmented. So there is a nat-
uralness in taking sentences as the basic bearers of the truth
values.

Tarski showed us how to think of truth as correspondence
to fact without having to take seriously a metaphysics of facts,
or a correspondence relation between sentences as wholes and
chunks of the world, for Tarski showed us how to get away with
just sequences of objects and the satisfaction relation.\textsuperscript{3} For our

\textsuperscript{2} This phrase is John Austin's from his "Truth" reprinted in his Philosophical Papers, eds. Urmson and Warnock, Oxford, 1961.

\textsuperscript{3} For a hint of what is going on here, consider the subject-predicate sentence "Socrates is wise". Plato might have the subject "Socrates" de-
ote note Socrates, the predicate "is wise" denote wisdom, and the sentence
true if and only if the denotation of its subject has the property denoted
by its predicate. Tarski cuts out the property. So while "Socrates" still de-
present purposes, no logical technicalities matter much. Consider instead the following simple picture:

![Diagram showing the relationship between language, reference, and the world.]

Truth is a matter of relations between words and the world. The sentence "The cat is on the mat" will not be true unless out there in the world there is a flesh and blood cat denoted by the singular term "the cat" that is the subject of the sentence and unless there is a mat denoted by the singular term "the mat" that is the object of the preposition "on". Of course, the cat must also be on the mat, and in his middle period at least, the historical Plato was much concerned to argue that truth here requires that the predicate "is on" of our sentence must also denote. But, whether predicates refer is not an issue that will concern us here. Instead, for us, the slogan that truth is correspondence to fact boils down to saying that truth requires reference to objects. It is singular terms that denote, but along with the proper names, definite descriptions and demonstratives usually counted as singular terms, we will also include the variables of quantification; so we are assimilating objects.

notes Socrates, an object satisfies "is wise" if and only it is wise, and the subject-predicate sentence is true if and only if the denotation of the subject satisfies the predicate. For more detail, see Mark Platts, *Ways of Meaning*, London: Routledge and Kegan Paul, 1979.
as values of variables to objects as denotata of singular terms more conventionally conceived. The idea of the first premiss of the metaphysical horn of Benacerraf's dilemma is that no sentence of a developed body of sentences is true unless there are in the offing singular terms referring to objects; truth requires reference to objects.

As a gloss, we might here repeat a comment Hilary Putnam used to add in lectures a quarter of a century ago: there is not a shred of linguistic evidence that the predicate “is true” of sentences is ambiguous. A conjunction of any two conjuncts is true if and only if both conjuncts are true, and there is no equivocation in saying that it is true that the earth is round and there are infinitely many primes. It is a single phenomenon of truth that emerges across the whole spectrum of knowledge.

The second premiss of the metaphysical horn of the dilemma is that mathematics is (at least by and large) a body of truths. As noted above, since ancient times mathematics has been held up as a paradigm of absolute truths without qualification, and to this day it would seem heroic, even quixotic, to deny it. Even those who at least verbally seem to deny it do not seek to drive mathematicians from the temples of the academy as if they were charlatans like astrologers.

It follows from these two premisses alone that there are objects, like numbers, functions and sets, denoted by the singular terms and variables of the truths of mathematics. This comes to no more than saying that the theorem in Euclid that there are infinitely many prime numbers is not true unless there are numbers. (Putnam used to quip that Quine's criterion of ontological commitment says that what you say there is is what you say there is.)

Although the claim that there are numbers already sticks in many a metaphysical craw, it is not yet quite the thesis of pla-

---

tonism stated above namely, that there are abstract objects. To 
draw that conclusion, we need a third premiss, that numbers 
are abstract. This is probably obvious, but it might be fun to try 
to marshall some arguments for it. Here are three:

(1) If the natural numbers were any old physical objects, so 
that, for example, zero is the moon, one is the asteroid Ceres, 
and so on, then it seems inexplicable that examining these 
physical objects plays no part in the justification of mathematical beliefs.

(2) For reasons we see under the epistemological horn of 
the dilemma, the favoured physical objects to be, or replace, the 
numbers are the numerals. If the numerals are really to be physical, they should be actual inscriptions (strips of molecules) 
or utterances (sound waves). But as Frege observed long ago 
against the formalists of his day, even if the human race lasts 
forever, at each point in our history we will only ever have in-
scribed or uttered finitely many numerals. So since there are 
ininitely many numbers, there just aren’t enough physical numerals to be all the numbers. If most numbers aren’t physical numerals, parity of reasoning or uniformity of kind urges that none are.

At this point it might occur to one that perhaps we could se-
cure enough numerals by construing them as possible inscrip-
tions or utterances or as sequences of physical inscriptions of, 
say, digits. But a sequence is a kind of set and thus abstract; 
and it seems clear that a merely possible physical object is not a physical object.

Frege’s argument may seem at first like mere trickery. But its 
deep point is that mathematics, like most sciences, has a sub-
ject matter, and that the subject matter of mathematics is the 
infinite, either because, like number theory, it treats of object 
of which there are infinitely many (and so needs proof instead

---

of examination of all cases), or else because, like analysis, it
treats of infinitely many infinite objects (like real numbers), or
ext else because, like set theory, its *raison d'être* is as a theory of
infinity *per se*. The infinitary character of the subject matter
of mathematics is a high hurdle for all attempts to render that
matter material.

(3) This last argument is probably too philosophical. It is an
ancient and honourable piece of philosophical lore that if there
are numbers, they exist necessarily. But at least since part IX of
Hume's *Dialogues Concerning Natural Religion*, we have been
pretty sure that no physical object exists necessarily, since for
any one physical object we can imagine what the rest of the
world would be like without it. Since most of us believe that
life evolved only late in the history of the universe, it is even
easier to believe that no mental entity exists necessarily either.
It follows that no number is either mental or physical. So, on
the usual three-way split that what there is is either mental,
physical or abstract, it follows that numbers are abstract.

To sum up, mathematics is a body of truths, like Euclid's
theorem that there are infinitely many prime numbers. Truth
requires reference to objects, so this theorem will not be true
unless there are numbers. Since numbers are abstract, there
are abstract objects. That is the metaphysical thesis of platon-
ism, so we have completed our exposition of the first horn of
Benacerraf's dilemma.

Let us take a brief intermission before looking at the second
horn, for someone surely wants to ask skeptically what it is to
be an abstract object. That is a good question. They are also
good questions what it is to be a physical object and what it is
to be an object. The distinction between the abstract and the
concrete may be too basic for it to be reasonable to expect an
enlightening analysis of it into a complex of simpler bits. So
instead it might be shrewd to inquire after the basic laws relat-
ing to the physical objects or the abstract objects. The science
of physics is our repository of our best shots at the basic laws
of matter. So it seems wise to send one asking what a physical object is to study physics. Maybe one day we will have good reason to believe in abstracta (like perhaps, propositions or properties) other than those like numbers, functions and sets studied in mathematics. But for now pure mathematics seems like our repository of our best shots at the laws of abstract. So it seems wise to send one, asking what an abstract object is to study mathematics; mathematics is the best statement of platonism. As for objects in general, logic (first order quantification theory with identity) seems our repository for our best shots at stating the laws to which any old objects are subject, so one asking what an object is might be sent to study first year logic. This last is a point long urged by Frege, Russell and Quine.

But maybe there are a few things, mostly negative, that we can say about very abstract objects like numbers, functions and sets. They are not located in space. That is the point of the adverb “very”. Consider geometrical objects. Take the present centres or mass of the earth and the moon. They determine a unique straight segment joining them. Consider the sphere of radius one mile about the midpoint of that segment. That sphere is a geometrical object in space, but there need be no matter anywhere in it. So it seems right to think of it as at least more abstract than a galaxy or a glacier, if not abstract tout court. But since it is located in space, it also seems right to think of the sphere as less abstract than the number π, which lacks location; hence the description of numbers as very abstract. (There is an apparently idle dispute about whether numbers are eternal or atemporal. Forced, to choose it seems better to choose eternity, since there were infinitely many primes long before life evolved, and will be long after its extinction.)

For our purposes, perhaps the most important negative general feature of very abstract objects like numbers is that it seems as close to axiomatic as metaphysical claims get that very abstract objects like numbers are utterly, causally inert. They do not reflect light, nor do they bump into anything. The natu-
rationalism prevailing now among us has it that, in Hume's words, causation is the cement of the universe. One convinced of platonism should deny that our conventional cosmology is the whole story.

The second horn of Benacerraf's dilemma may be thought of as a sequence of three claims of increasing specificity rather than an argument. Suppose a person, A, knows that the cat is on the mat. How are we to understand this? The first of our three claims is that we are happiest in thinking we understand how A knows that the cat is on the mat only when we can at least sketch out some sort of transaction or commerce or connection between A and the objects, the cat and the mat, reference to which is required for the truth of the sentence stating A's belief. The idea is that, given that truth requires reference to objects, knowledge of truth is best understood in terms of some transaction, commerce or connection between the knower of the truth and the objects required for truth which justifies the knower's belief about them.

This first claim is slack. Two of its central terms "transaction" and "commerce", are metaphors, and "connection" seems very general. Our second claim is a sharpening of our first. It is that the only generally agreed basic mode of commerce between people and objects that justifies true belief, about them is perception. Even Plato seeking to account for knowledge of the forms in the analogy between the sun and the form of the good in the allegory of the cave, or in describing how the slaved boy got his geometrical knowledge to be later recollected, seems clearly to hew to a perceptual model of knowledge acquisition; and the same seems implicit in rationalist metaphors about the light of reason. It is not so much that of two rival theories of how knowledge of objects is to be understood, empiricism and rationalism, the first accounts for the data better than the second. It is rather that rationalism never made out a connection between people and objects alternative to people's perception of them and which nonetheless justifies people's true
beliefs about them. To put it badly, rationalism does not exist; we are all, even Plato and Descartes, empiricists in our bones.

Our third claim is a sharpening of our second due to Paul Grice. 6 To present it, we back up a bit and consider perception for a while. Macbeth’s hallucination of a dagger could have been so thorough that there be no way for him to tell whether he is hallucinating or is in a case like that of a man, call him Fred, seeing a real dagger with its handle toward his hand hanging in the air before him (perhaps suspended between concealed magnets). The traditional argument from illusion infers from this that there is visual experience common to seeing and visual hallucination, and John Austin’s bald denial 7 hardly renders the inference shaky. But because visual experience is common to sight and visual hallucination (and visualizing in the mind’s eye and, probably, dreaming), visual experience is not all there is to vision. At this point the tradition adds veridicality. Veridicality is to be to experience as truth is to thought. As a thought is true when the world is as the thinker thinks it to be, so visual experience is veridical when the world is as it looks to the person having the visual experience. Note that both truth and veridicality require no more than bare conjunction: a thought is true when the world is a certain way and the thinker thinks it to be that way; visual experience is veridical when the world is a certain way and that is how it looks to the person having the visual experience. This is where Grice comes in by objecting that bare veridical visual experience is not enough for sight.

Grice objects by example. Imagine a man seated in a darkened room. He is having a visual experience as of a lighted candle about a foot tall a few feet in front of him. In fact there is a lighted candle about a foot tall a few feet in front of him.


So if bare veridical visual experience were enough for sight, the man would be seeing the candle in front of him. But, we may also suppose, no light reaches the man from the candle in front him because between them is an opaque object blocking off the light from the candle. So why is he having his visual experience? Because, we may suppose, a long way off to his right there is a lighted candle about a foot tall and the start of an ingenious sequence of mirrors, the last of which is the opaque object in front of him, reflecting light from the distant candle to cause his visual experience. Suppose, Grice asks, you had to pick one of these two candles and say that it is the candle the man is seeing; which would it be? Everyone seems to have an opinion on this question, and the same opinion; forced to pick, one says it is the distant candle from which comes the light causing his visual experience. So, putting the pieces together, Grice claims that veridical visual experience is sight only if it is caused by that in virtue of which it is veridical. In other words, two relations between visual experience and objects are required for sight: veridicality, in that the objects should be as they look to the person with the visual experience to be; and causation, in that how the objects are should cause the person's visual experience. Veridical visual experience is vision only if it is caused by that in virtue of which it is veridical. So, we may conclude, perception is by nature a causal process.

At this point, we should add two remarks. Benacerraf did not, as we have, appeal to Grice and the causal nature of perception. Instead Benacerraf appealed to Goldman's causal analysis of the meaning of (certain uses of) the verb "to know,"8 where Goldman's analysis was a response to Gettier's famous counter-examples to the sufficiency of justified true belief for

knowledge. We prefer our route through Grice and the nature of perception because it seems to go more to the heart of the matter, and because it avoids claims about meaning parochial to an earlier phase of philosophy. But Grice and Goldman can both be ranged under a philosophical project of naturalizing the mind. Assuming the conventional cosmology that causation is the cement of the universe, the project is to naturalize the mind by fitting it into the causal nexus nature is taken to be. Under this project were presented causal accounts not just of perception and knowledge, but also of action, de re belief, and naming. Second, we should remark in fairness to Grice that he was perfectly clear that not just any old causation of veridical visual experience by that in virtue of which it is veridical suffices for sight. Suppose, for example, you are captured by a mad scientist who subjects you to total sensory deprivation and then injects you with chemical C. This injection causes in you only one psychological effect, a visual experience of the mad scientist injecting you with chemical C. Here we have veridical visual experience caused by that in virtue of which it is veridical, and yet the consensus is that in this case you do not see the scientist inject you with chemical C. Granted the consensus, only a special kind of causation of veridical visual experience by that in virtue of which it is veridical suffices for sight. Grice labeled this kind appropriate, and tried to foist the specification of appropriateness off onto natural scientists. This seems a very speedy abdication of intellectual responsibility. The problem of appropriate causation seems general in that most of the causal


12 Saul Kripke, Naming and Necessity, Blackwell, 1980.
accounts of mental functions mentioned above are subject to counter-examples, like the mad scientist and chemical C, that seem to call for restriction of the required causation to some appropriate sort.\textsuperscript{13}

But to exposite the second horn of Benacerraf’s dilemma, all we need is Grice’s result that causation is necessary for perception, not a specification of which sort of causation suffices. For if knowledge is best understood in terms of a transaction between the knower and the objects in virtue of which what he knows is true, if the only such mode of transaction is perception, and if perception is by nature causal, then it is at least obscure how a person could have any knowledge of a subject matter that is utterly inert, and thus with which he could have no causal commerce. And yet by the first horn of the dilemma, the numbers, functions and sets have to be there for the pure mathematics of numbers, functions and sets to be true. Since these objects are very abstract, they are utterly inert. So it is at least obscure how a person could have any knowledge of the subject matter needed for the truth of the pure mathematics of numbers, functions and sets. As promised, Benacerraf’s dilemma is that what seems necessary for mathematical truth also seems to make mathematical knowledge impossible.

Benacerraf uses the hypothesis that his dilemma is the basic problem addressed in the philosophy of mathematics to explain a striking feature illustrated by some textbooks on that subject earlier in our century. Once upon a time it seemed as if one could be a logicist, a formalist or an intuitionist, but that these and only these were the positions available. But why? Granted Benacerraf’s dilemma, one can at best preserve intact only one of its horns, and if so, one can solve the dilemma only by somehow blunting the other. So, it seems, either one sticks

\textsuperscript{13} For more on the problem of appropriate causation, see Cristopher Peacocke’s discussion of it in Holistic Explanation, Oxford, 1979, the references he cites, and W.D. Hart, The Engines of the Soul, Cambridge, 1988, pp. 56–58.
with the platonism required by the metaphysical horn and modifies one’s epistemology, or else one sticks with the empiricism required by the epistemological horn and modifies one’s metaphysics. The logicism of Frege and Russell was a species of platonism. On the epistemic side, Frege talks of grasping abstracta, Russell, of acquaintance with universals. And Gödel of our having something like perception of the objects of set theory. Each is strikingly taciturn about the epistemic faculty he claims, but it seems plausible that each is cutting his epistemological coat to suit his metaphysical cloth.

We mentioned above the usual three way split of what there is into mental, physical and abstract. Many people claim to find abstract objects weird, and the epistemological horn of Benacerraf’s dilemma is one good way to back up that inductive with thought. If one rejects the platonism of the metaphysical horn, one is left with a two way split between the physical and the mental (where many nowadays identify the second with the first, of course). Those who favour the physical reap the epistemological harvest that (many) physical objects are accessible to ordinary sense perception. But then they owe us an account of mathematical truth without platonism. Well, what we see and hear when we learn mathematics are inscriptions and utterances. So formalism, which is what nominalism in philosophy of mathematics is called, is motivated epistemologically. On the other hand, we at least used to take introspection and Kant’s inner sense serious as the faculties in the exercise of which we come to know our own minds. Granted the epistemological legitimacy of inner sense, Brouwer’s positive account of mathematical assertions as reports of mental phenomena in the speaker is also epistemologically motivated.

14 See, for example The Problems of Philosophy, Oxford, 1959, p. 103
The textbook triple option (logicism, formalism, intuitionism) reflects the metaphysical catalogue of abstract, physical and mental entities, where the platonist has an easy way with truth but a hard one with knowledge, while the formalist and intuitionist have a more comfortable knowledge but a more awkward time securing objective truth for received mathematics (and the intuitionist gives up the effort). It confirms Benacerraf's hypothesis that it thus explains the striking triple option of the textbook.

Benacerraf's dilemma is an antinomy at the confluence of metaphysics and epistemology. There are other problems here that from a sufficiently abstract point of view, are analogous to Benacerraf's dilemma. Modal claims are claims about what is (absolutely) necessary and what is (merely, i.e., without being actual) possible. For example, the thesis of cartesian dualism is the modal claim that you (a mind, a self) do not depend for your existence on that of your body and that you could be disembodied. Years of ordinary language philosophy notwithstanding, the mind-body problem seems not a pseudo-problem but a genuine one. So either the thesis of dualism is objectively true or it is objectively false. The objectivity of a truth is its being true independently of whether we say, think or believe it true, and answering to objects that are as they are independently of how we say, think or believe them to be seems the only way truth is objective. So on the metaphysical side, objectivity of modal truth seems to push us toward possible worlds independent of us. But mere possibilities would be just as inert causally as actual very abstract objects. So as in Benacerraf's dilemma, what seems necessary for objective modal truth seems to make modal knowledge impossible; there are no telescopes, not even the imagination, that let us see what is going on in other possible worlds.16

An analogous antinomy may underlie philosophy of mind, the mind-body problem roughly symptomatic of its metaphysical horn, and the problem of other minds, of its epistemological horn. In the mind-body problem, on which professional specialists have focused for the last twenty or so years, dualism seems the natural metaphysics, despite the popularity of materialism among the specialists. The metaphysical argument here is Descartes': as a matter of the epistemology of modality, what one can imagine is possible; and one can imagine being disembodied; so one could be disembodied, which is the thesis of dualism. But dualism makes the epistemic problem of other minds especially acute, for while one can see other bodies and their activities, it is a version of the inter-action problem for dualism that it is at least obscure whether one could perceive other minds not dependent for their existence on the bodies in which they might be lodged. Skinner then whispers that were your mother’s love for you part of her behaviour, you could know by ordinary sense perception that she loves you. So behaviourism is the natural empiricist epistemology of other minds. The problem, the reason for philosophy of mind, is that dualism and behaviourism are inconsistent.

The natural metaphysics of space is Newton’s. For suppose with him there were just two spheres of equal mass joined by a rope. Were they rotating about their common centre of gravity, there would be tension in the rope, while it would be slack if they were at rest. This is convincing, Mach notwithstanding, even though there are no other material bodies with respect to which they are at rest or in motion. That way lies absolute space as the arena with respect to which they are in motion or at rest. Leibniz then objects that were the separation between bodies constant but the system of bodies moving at a constant five miles an hour to the left in absolute space, this motion would be utterly undetectable. So for motion with respect to space to be knowable, space should be reducible to relations between bodies. The motive for relational theories of space is
epistemological. The problem, a reason for philosophy of space, time and spacetime, is that space cannot be both absolute and relational.

We do very serious things to people in the name of morals, so we want our morals to be more than our pleasures or whims; we want an objectivity for our values, an independence of them from what we take them to be. That way lies Plato’s Form of the Good and Moore’s non-natural property of goodness. But a plain problem with either is epistemological. Without somehow naturalizing them, how can we make sense of moral knowledge? Utilitarians and subjectivists then whisper that were the good what maximizes happiness, goodnes would be open to ordinary empirical investigation in psychology (or perhaps sociology). The problems and a reason for moral philosophy, is that the good as happiness seems incompatible with that objectivity and independence of our wishes that we wanted for value.17

In Benacerraf’s dilemma and its analogues, the second horn is epistemological. There the point is always that empiricism, which counts perception as the basic mode of justification, is our favoured epistemology, that perception is a causal process, and yet that the objects required by the metaphysical horn, seem utterly inert. While these are epistemological points, they have their metaphysical dimensions. For the epistemological difficulty is also a failure to fit knowing minds and known objects together in a universe of which the cement is causation. Could it be the conventional cosmology that is at the root of these problems?18

17 Ana Maria Richter points out a version of this analogue in the philosophy of law. The legal positivist (Austin, Kelvin, Hart) who claims that law in nothing but statute and precedent has an obvious epistemic advantage because what is written is visible. But where statute and precedent give out and judges are expected to exercise discretion, a notion like Dworkin’s of judges as articulating conceptions seems morally preferable to the arbitrariness to which legal positivists seem here reduced

18 For more on this theme, see W. D. Hart, “Natural Numbers”, the next issue of Crítica.
RESUMEN

Un problema básico que considera la filosofía de la matemática se formula de la siguiente manera. Del lado metafísico, la verdad requiere que se haga referencia a objetos: así, puesto que es verdad que hay infinitamente muchos números primos, hay infinitamente muchos números primos. Así pues, hay números, que son objetos abstractos y, por esto, completamente inertes. Pero, del lado epistemológico, el conocimiento se entiende mejor en términos de una transacción entre el sujeto cognoscente y los objetos por virtud de los cuales es verdadero lo que él conoce. El único modo aceptado de tal transacción es la percepción y, conforme a la argumentación de Grice, la percepción es causal por naturaleza. Así pues, por una parte, la verdad matemática requiere que haya objetos de los que estamos causalmente aislados mientras que, por otra parte, el conocimiento en general y, por esto, el conocimiento matemático en particular, requiere que haya objetos a los que tengamos acceso causal. En breve, lo que parece necesario para la verdad matemática, hace imposible el conocimiento matemático.

[Traducción de José Antonio Robles]