THOUGHT EXPERIMENTS: A PLATONIC ACCOUNT[1]

James Robert Brown

MATHEMATICS *VS* PHYSICS?

METAPHORS about the book of nature being written in the language of mathematics suggest a very intimate connection between mathematics and physics. But few modern philosophers, if any, hold this as their considered opinion. I suspect that the vast majority hold a view something like this: Mathematics is a theory of formal systems; it has nothing to do with the physical world at all, though problems in the physical sciences may suggest lines of mathematical research. The results of mathematics are known entirely a priori by deriving them from axioms which are taken to be true by convention, or true in virtue of the meanings of the terms involved, etc.

By contrast, the common picture of physics and the other natural sciences is something like this: Physics is an attempt to describe the physical world. Mathematics provides useful tools in the sense that theorizing about the physical world involves use of mathematical models, and these mathematical models are given a priori. But our knowledge that the physical world conforms to some particular mathematical structure is itself a posteriori and so, unlike mathematics, it is fallible.

To a large extent much of this is summed up in a oft-cited remark of Einstein's: "... as far as the propositions of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality."[2] Einstein goes on to explain why he holds this view. I suspect his reasons are pretty universal.

> ... complete clarity as to this state of things became common property only through that trend in mathematics which is known by the name "axiomatics." The progress achieved by axiomatics consists in its having neatly separated the logical-formal from its objective or intuitive content; according to axiomatics the logical-formal alone forms the subject matter of mathematics, which is not concerned with intuitive or other content associated with the logical-formal.

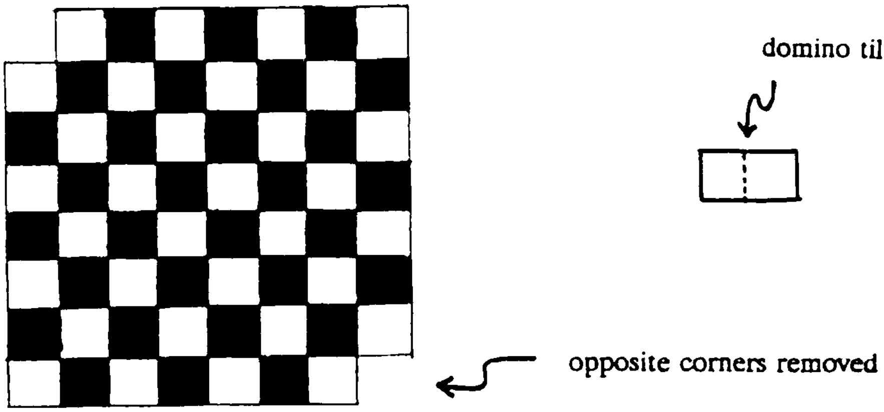
In brief, the methods and the ontology of mathematics are entirely different than the methods and ontology of physics; physics is empirical and about the material world while mathematics is neither. Even more briefly, mathematics and physics are utterly disunified.

MATHEMATICAL INSIGHT

In a very influential paper in the philosophy of mathematics, Paul Benacerraf[3] sets up a dichotomy and a dilemma. He sees two main approaches to mathematics: the combinatorial and the semantic. The view sketched out above is combinatorial, while platonism is the paradigm example of a semantic account. Benacerraf's dilemma is this: the virtues of one account are the sins of the other. For example, platonism is very good at making sense of the idea of mathematical truth—mathematical propositions are true in the same way that the propositions of physics are. By contrast, any sort of combinatorial account will have to link truth with provability, which will turn out to be highly problematic for a number of reasons. (For example, Gödel's theorem shows clearly that truth and proof are distinct notions.) On the other hand, combinatorial accounts do great justice to our knowledge—we know what we know because we have a proof. Platonism, by contrast, appears to embrace some very mysterious processes to account for how we humans who are inside space and time can come to know anything about the abstract entities outside of space and time. So, we can have an acceptable epistemology or an acceptable semantics, according to Benacerraf—but not both.

By "proof" in his combinatorial account, Benacerraf means a derivation. That is, we have a proof of P when we derive it from some set of accepted axioms. So when we prove P we are really showing that "$\vdash$ Axioms $\supset$ P " is a logical truth. I do not doubt for a moment that such a logical relation holds, but could Benacerraf really be right in thinking that this derivation relation is the ground for our *knowledge* that P?

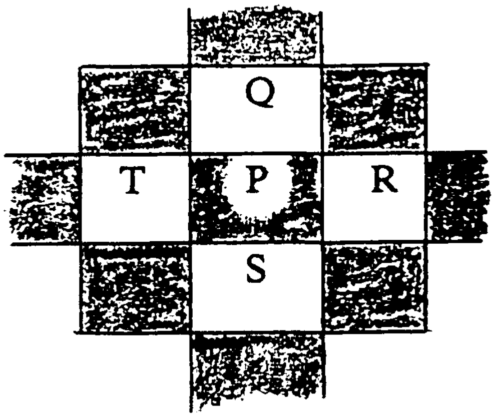
Consider this simple example. We are given a chessboard with two opposite corners removed. We are also give an unlimited number of domino tiles which each cover two squares of the board. Problem: can the board be tiled with the dominos so that each square is covered and there is no overlap?



The answer is no. I will give two proofs.

The first proof is not in any way combinatorial. It runs as follows: There can be no tiling since each tile covers one coloured and one non-coloured square. Thus, a complete tiling requires an equal number of each, but we do not have it in this board. So, tiling the board is impossible.

Here is a second proof which is combinatorial. Label each square P, Q, R, etc., meaning "P is covered", "Q is covered," and so on.



A tile which covers P also covers one and only one tile beside P. A full description of the tiling around P is:

$$P\&Q \vee P\&R \vee P\&S \vee P\&T\ \&$$
$$P\&Q \rightarrow \sim(P\&R \vee P\&S \vee P\&T)\ \&$$
$$P\&R \rightarrow \sim(P\&Q \vee P\&S \vee P\&T)\ \&$$
$$P\&S \rightarrow \sim(P\&Q \vee P\&R \vee P\&T)\ \&$$
$$P\&T \rightarrow \sim(P\&Q \vee P\&R \vee P\&S)$$

This is the description *for just one square.* The full description for the entire chessboard will be about sixty-four times as long. The conjunction of all these propositions is a contradiction. Thus, its denial is a tautology, and hence, derivable in any complete system of propositional logic.

But does anyone believe the tiling theorem because of this proof? I doubt it. And certainly no one came to first know the result by actually deriving it in a system of propositional logic. It is the short—non-combinatorial—proof that makes us believe the tiling theorem.

Just in case the ordinary chessboard example is not convincing, notice how much worse the combinatorial explosion becomes as we increase the size of the board. Yet the short proof works equally well for any $n \times n$ board (with two opposite corners removed) no matter how big n is.

This example brings us back to the platonist side of Benacerraf's dichotomy. Our knowledge of the tiling theorem is not based on a derivation. The combinatorial account not only fails to do justice to *truth,* but comes up short on *knowledge* as well. Rather, our knowledge is based on a grasp of some sort of pattern. There is some sort of relationship which exists objectively, independently of us, and we can somehow see it. Roger Penrose puts it vividly: "...whenever the mind perceives a mathematical idea, it makes contact with Plato's world of mathematical concepts... When one 'sees' a mathematical truth, one's consciousness breaks through into this world of ideas, and makes direct contact with it..."[4]

There are several ingredients involved in platonism. (1) There are abstract objects existing outside of space and time. (2) The way these objects are is what makes our mathematical statements true or false. (3) The mind can grasp or intuit (some of) them. (4) Our mathematical knowledge is a priori in the sense of being independent of the physical senses; but it need not be infallible.

The account sketched is similar to Gödel's, the foremost platonist of recent times, who writes

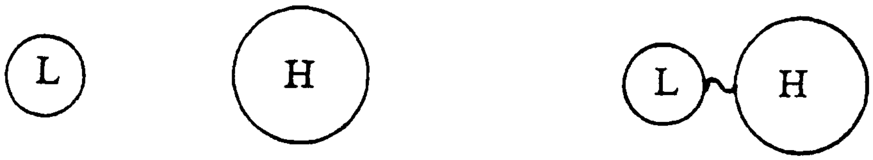
> ... despite their remoteness from sense experience, we do have something like a perception also of the objects of set theory, as is seen from the fact that the axioms force themselves upon us as being true. I don't see any reason why we should have any less confidence in this kind of perception, i.e., in mathematical intuition, than in sense perception.[5]
>
> ... the assumption of such objects is quite as legitimate as the assumption of physical bodies and there is quite as much reason to believe in their existence. They are in the same sense necessary to obtain a satisfactory system of mathematics as physical bodies are necessary for a satisfactory theory of our sense perceptions...[6]

Of course, the platonism I am sketching faces the same objections made against Gödel; namely, how is such a "grasping" or "perception" possible, since abstract objects, being outside of space and time, cannot causally interact with us? Meeting this objection largely amounts to showing that the causal theory of knowledge is wrong. Since I have attempted that elsewhere,[7] I will pass on to my main topic.

Thought Experiments

Thought experiments are performed in the laboratory of the mind. Beyond that bit of metaphor it's hard to say just what they are. We recognize them when we see them: they are visualizable; they involve mental manipulations; they are not the mere consequence of a theory-based calculation; they are often (but not always) impossible to implement as real experiments, either because we lack the relevant technology or because they are simply impossible in principle.

Let us now look at the finest example of a thought experiment ever: Galileo's wonderful argument in the *Discoursi* to show that all bodies, regardless of their weight, fall at the same speed.[8] It begins by noting Aristotle's view that heavier bodies fall faster than light ones (H > L). We are then asked to imagine that a heavy cannon ball is attached to a light musket ball. What would happen if they were released together?



Reasoning in the Aristotelian manner leads to an absurd conclusion. First, the light ball will slow up the heavy one (acting as a kind of drag), so the speed of the combined system would be slower than the speed of the heavy ball falling alone (H > H+L). One the other hand, the combined system is heavier than the heavy ball alone, so it should fall faster (H+L > H). We now have the absurd consequence that the heavy ball is both faster and slower than the even heavier combined system. Thus, the Aristotelian theory of falling bodies is destroyed.

But the question remains, "Which falls faster?" The right answer is now plain as day. The paradox is resolved by making them equal; they all fall at the same speed (H = L = H+L).

Not all thought experiments work this way. Some only destroy. I call this class of thought experiments *destructive*. Typically, the destructive thought experiment is some sort of *reductio ad absurdum* of a pre-existing theory.

Young Einstein imagined himself chasing a light beam, and he wondered what it might look like, if he could run as fast. A light beam is an oscillating electromagnetic field. The magnetic field exists because of a changing electric field and the electric field exists because of the changing magnetic field—a kind of leap-frog effect where the frogs exist only so long as leaping. But if Einstein ran along side of such a wave, it would appear stationary, in which case it could not exist at all. This thought experiment is a beautiful *reduction ad absurdum* of the joint theory of classical mechanics and Maxwell's electrodynamics.

Schrödinger's cat is yet another example of a purely destructive thought experiment. In this case Schrödinger's target was the Copenhagen interpretation of quantum mechanics. Of course, this thought experiment did not show that the Copenhagen interpretation is logically inconsistent, as so many destructive thought experiments typically do to their targets, but rather shows that the Copenhagen interpretation is in flagrant violation of well-entrenched common sense.

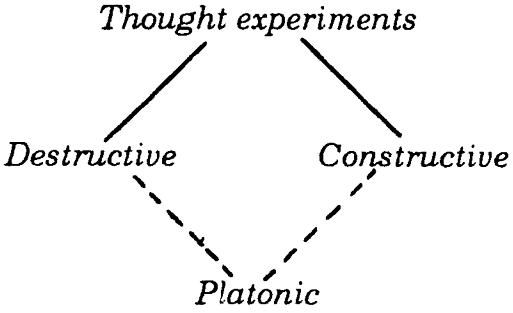
Many thought experiments fall into a completely different class. Instead of playing a refuting role, they provide supporting evidence for a theory. I call this class *constructive*.[9]

Newton imagined a bucket of water in an otherwise empty space. In one state the water is at rest with respect to the bucket and its surface is flat; in a second state there is relative motion of the water and the bucket; in a third state the water and bucket are again at rest with respect to one another, but this time the water's surface is concave. What is the difference between states one and three? Newton offered the following explanation: in the third case, but not the first, the water and the bucket are in motion with respect to absolute space. Therefore, absolute space exists.

Maxwell was a champion of the statistical-kinetic theory of heat, but he had to face the problem that he could not derive the second law of thermodynamics in any strict form, but only a probabilistic version. He could only conclude that it is very unlikely, for example, that heat could pass from a cold body to a hot one. To some, any possibility at all of this happening smacked of absurdity. Maxwell's demon thought experiment showed that it is indeed possible that heat could pass from a cold to a hot body. The demon controls a door between two adjacent chambers, one hot, the other cold. The average speed of the molecules is different in the two chambers; there are, however, some fast ones in the cold room. These the demon lets into the hot room while letting slow molecules from the hot room into the cold. In this way heat has passed from the cold to the hot chamber, since the average speed of the molecules has increased in the hot room and decreased in the cold.

As a final example of a constructive thought experiment consider Einstein's elevator. The observer is completely enclosed inside an elevator, but cannot tell whether the elevator is stationary in a gravitational field, or is accelerating upward in force-free space. The principle of equivalence says there is no difference between these two states. Suppose the elevator is actually accelerating and that a beam of light enters through a small hole in one wall. The beam will take a finite time to cross the elevator, during which the elevator moves up. Thus, to the observer, the beam will appear to be bent downward. Consequently, by the principle of equivalence, the beam of light would be bent in a gravitational field, as well.

Physics is full of thought experiments which are either destructive or constructive. But wonderful though they are, there is a still more remarkable class of examples. I call this final class *platonic*. These thought experiments are simultaneously constructive and destructive. They destroy the old and they create the new in a single blow.



There are few examples of this remarkable type of thought experiment. I think the argument of Einstein, Podolsky, and Rosen is a candidate. This thought experiment was both a *reductio ad absurdum* of the Copenhagen view *and* a positive argument for local hidden variables. (At this point I should stress the fallible nature of thought experiments. The fact that local hidden variables are now known to be hopeless does not affect the analysis of how the thought experiment works.) Yet another possible candidate is Leibniz's argument for *vis viva*. This simultaneously destroyed the Cartesian view of what is conserved *and* established a new account. But the greatest example of all is Galileo's wonderful thought experiment on free fall which I sketched above. For purposes of illustration I will stick to it.

Though the reasoning in Galileo's thought experiment is fallible, it is *not* a piece of standard, empirical, conjectural, a posteriori knowledge. Rather, we are justified in calling this a case of a priori science. Here's why: First, *there have been no new empirical data.* I suppose this is almost true by definition; being a *thought* experiment rules out new empirical input. It's not that there are no empirical data involved in the thought experiment at all. The emphasis is on *new* sensory input; it is this that is lacking in the thought experiment. What we are trying to explain is the *transition* from the old to the new theory and that is not readily explained in terms of empirical input unless there is new empirical input.

Second, *Galileo's new theory is not logically deduced from old data. Nor is it any kind of logical truth.* A second way of making new discoveries—a way which does not trouble empiricists—is by deducing them from old data. Perhaps Galileo's thought experiment is really an argument. Is this plausible? I think not. The premisses of such an argument could include all the data that went into Aristotle's theory. From this Galileo derived a contradiction. (So far, so good; we have a straightforward argument to this point.) But can we derive Galileo's theory that all bodies fall at the same rate from these same premisses? Well, in one sense, yes, since we can derive anything from a contradiction; but this hardly seems fair.[10] What's more, whatever we can derive from these premisses is immediately questionable since, on the basis of the contradiction, we now rightly consider our belief in the premisses to be undermined.

Might Galileo's theory be true by logic alone? To see that the theory that all bodies fall at the same rate is not a logical truth, it suffices to note that bodies might fall with different speeds depending on their colours or on their chemical composition as has recently been claimed.[11] These considerations undermine the argument view of thought experiments.[12]

Third, *the transition from Aristotle's to Galileo's theory is not just a case of making the simplest overall adjustment to the old theory.* It may well be the case that the transition was the simplest, but that was not the reason for making it. (I doubt that simplicity or other aesthetic considerations ever play a useful role in science, but for the sake of the argument, let's allow that they could.) Suppose the degree of rational belief in Aristotle's theory of falling bodies is r, where $0 < r < 1$. After the thought experiment has been performed and the new theory adopted, the degree of rational belief in Galileo's theory is r', where $0 < r < r' < 1$. That is, I make the historical claim that the degree of rational belief in Galileo's theory was *higher* just after the thought experiment than it was in Aristotle's just before. (Note the times of appraisal here. Obviously the degree of rational belief in Aristotle's theory *after* the contradiction is found approaches zero.) Appeals to the notion of smallest belief revision won't even begin to explain this fact. We have not just a new theory—we have a better one.

As well as these there are other reasons which suggest the example yielded a priori knowledge[13] of nature, but possibly the most interesting and most speculative has to do with its possible connection to the realist account of laws of nature recently proposed by Armstrong, Dretske, and Tooley.[14]

The new view stems from a discontentment with empiricist and nominalist regularity theories. Hume and his modern followers hold that a law of nature is a regularity. "It is a law that ravens are black" is analyzed as having the form $(\supset x)(Rx\supset Bx)$. Of course, it is immediately recognized that this will not quite do, since there are lots of regularities which are surely not laws of nature. "All the people in this room have silver coins in their pockets" fits the pattern, but is hardly a law. Regularity theorists such as Ayer, Braithwaite, and numerous others add a second condition to mark the difference. To be a law a statement first, must be a true universal statement, and second, must play a central role in our conceptual scheme. The coins-in-their-pockets example fails on this latter score.

The most obvious objections have to do with the utter subjectivity of the regularity view of laws. To be a law on the empiricist account something must be so recognized; consequently, there were no laws before there were people. Different people may have different laws simply because they place different (true) universal statements at the core of their webs of belief.[15]

The realist alternative is much more in harmony with our common sense beliefs in that laws are said to exist independently of us. But it does involve a commitment to abstract entities, that is, to universals. Along with individual ravens and black things there are abstract universals, ravenhood, *R*, and blackness, *B*, and they are in a relation of natural necessitation, *N(R,B)*. This relation between the universals entails the universally quantified sentence which expresses the regularity, but is not entailed by it. We have

$$\mathrm{N(R,B)}\rightarrow(\supset x)(\mathrm{R}x\supset \mathrm{B}x) \text{ and yet, } (\supseteq x)(\mathrm{R}x\supset \mathrm{B}x) \not\rightarrow N(R,B)\ .$$

According to Armstrong, Dretske, and Tooley this view of the laws of nature is pure metaphysics. It gives an account of the nature of laws and explains the regularities which obtain. But the way we learn about laws is the same as it was for the staunch empiricists: we look at individual instances of the regularity; we see ravens, never ravenhood.

I want to add an epistemological aspect to this metaphysical account of laws. My suggestion is simply this: in some thought experiments *we see the relevant laws*—not the regularities, but the universals themselves. Given the truth of platonism in mathematics (i.e., we can see some abstract entities), and given that Armstrong, *et. al.* have hit on the right account of laws of nature (i.e., laws are abstract entities), should we not expect to actually get a glimpse of them? In fact, would it not be a bit surprising if we were not able to somehow or other perceive at least some of these laws?

In conclusion, let me briefly contrast this form of platonism about physics with earlier forms of rationalism. Unlike Plato, Descartes, or Leibniz, etc., a priori knowledge on my view is neither certain nor innate. It is not put there by God; it is not remembered; nor is it infallible. But like the traditional rationalists, I hold that the abstract realm is perfectly real and that we can know something about it. Also unlike traditional rationalists, my view that we do sometimes acquire a priori knowledge of the physical world is not itself an a priori account. Descartes, by contrast, gave an a priori argument for his view that our knowledge is a priori. (It was clear and distinct to him that our clear and distinct ideas must be true.) The platonism I am defending is a conjecture—I simply claim that this hypothesis is the best explanation for the remarkable phenomena of (some) thought experiments. This is a style of arguing that most contemporary philosophers find congenial. Though I realize that the hypothesis itself is anything but.

NOTES

[1] Some of this essay draws on my forthcoming book on the subject. For a fuller discussion of thought experiments with lots of examples see *The Laboratory of the Mind: Thought Experiments in the Natural Sciences* (Routledge, 1991).

[2] Albert Einstein, "Geometry and Experience," reprinted in *Ideas and Opinions*, (New York: Bonanza Books, 1954), p. 233.

[3] Paul Benacerraf, "Mathematical Truth," reprinted in Benacerraf and Putnam (eds.), *The Philosophy of Mathematics*, 2nd ed. (Cambridge, 1983).

[4] Roger Penrose, *The Emperor's New Mind* (Oxford, 1989), p. 428.

[5] Kurt Gödel, "What is Cantor's Continuum Problem?", reprinted in Benacerraf and Putnam (eds.), *Philosophy of Mathematics* (Cambridge, 1983), p. 484.

[6] Kurt Gödel,"Russell's Mathematical Logic," reprinted in Benacerraf and Putnam (eds.), *Philosophy of Mathematics* (Cambridge, 1983), p. 456f.

[7] See J. R. Brown, "π in the Sky," in A. Irvine (ed.), *Physicalism in Mathematics* (Kluwer, 1989), and *The Laboratory of the Mind: Thought Experiments in the Natural Sciences.*

[8] Galileo, *Discourse on Two New Sciences* (trans. by S. Drake), (University of Wisconsin Press, 1974), p. 66f.

[9] Many variations within the constructive type of thought experiment are possible, but I will not pursue them here. A fuller taxonomy is given in *The Laboratory of the Mind: Thought Experiments in Natural Sciences.*

[10] Otherwise I could perform the thought experiment now and derive "The moon is made of green cheese."

[11] For example by Fischbach *et. al.* "Reanalysis of the Eötvos Experiment," *Physical Review Letters* (Jan. 1986). I suspect that the reason that Galileo's thought experiment works for light/heavy but not for colors is that the former are additive or extensive while the latter are not (i.e., combining two red objects will not make an object twice as red).

[12] I have been rather quick and perhaps even unfair in my dismissal of the argument view of thought experiments. John Norton (this volume) presents the argument view with great plausibility. I consider his views with more care in *The Laboratory of the Mind: Thought Experiments in the Natural Sciences.*

[13] The term "knowledge" may be too strong as it implies *truth*; "rational belief" might be better since, on my view, what is a priori could be false.

[14] See David Armstrong, *What is a Law of Nature?* (Cambridge, 1983); Fred Dretske, "Laws of Nature," *Philosophy of Science*, 1977; and Michael Tooley, "The Nature of Laws," *Canadian Journal of Philosophy*, 1977. Bas van Fraassen, *Laws and Symmetry* (Oxford, 1989), contains many important criticisms of the realist account of laws, but I will not try to defend Armstrong, *et. al.*, here.

[15] For detailed criticisms of the empiricists views of laws see the early chapters of Armstrong, *What is a Law of Nature*? or Tooley, *Causation: A Realist Approach* (Cambridge 1988).