The Bulletin of Symbolic Logic
Volume 4, Number 1, March 1998

AN EDITOR RECALLS SOME HOPELESS PAPERS

WILFRID HODGES

§1. Introduction. I dedicate this essay to the two-dozen-odd people whose refutations of Cantor's diagonal argument (I mean the one proving that the set of real numbers and the set of natural numbers have different cardinalities) have come to me either as referee or as editor in the last twenty years or so. Sadly these submissions were all quite unpublishable; I sent them back with what I hope were helpful comments. A few years ago it occurred to me to wonder why so many people devote so much energy to refuting this harmless little argument—what had it done to make them angry with it? So I started to keep notes of these papers, in the hope that some pattern would emerge.

These pages report the results. They might be useful for editors faced with similar problem papers, or even for the authors of the papers themselves. But the main message to reach me is that there are several points of basic elementary logic that we usually teach and explain very badly, or not at all.

In 1995 an engineer named William Dilworth, who had published a refutation of Cantor's argument in the Transactions of the Wisconsin Academy of Sciences, Arts and Letters, sued for libel a mathematician named Underwood Dudley who had called him a crank ([9] pp. 44f, 354). The case was dismissed. For myself I am more scared of the copyright law than the law of libel. After taking legal advice I decided not to quote any of the authors directly. The alternative was to write some letters saying in effect: 'I'm sorry we couldn't publish your paper as a contribution to logic. Can I please publish parts of it as examples of garbage?' Not much is lost, because almost all of the papers were written by manifest amateurs who had great difficulty explaining what they meant, and I freely admit that much of what follows is my own attempt to discern some thoughts behind the streams of words. This is in no sense a scientific analysis of experimental data.

§2. Cantor's proof. The authors of these papers—henceforth let me call them just the *authors*—seem to have read Cantor's argument in a variety of places. In my records only one author refers directly to Cantor's own argument [7]. One quotes Russell's 'Principles of mathematics' [20] later

Received Octobeer 6, 1997; revised October 15, 1997.

1079-8986/98/0401-0001/$2.60

in his discussion; he could have found the diagonal argument on page 365 of that book, in words that closely follow Cantor. Another cites Fraenkel 'Abstract set theory' [12] as his source, and another refers to Barrow 'Theories of everything' [2]. One contents himself with references to two earlier unpublished papers of his own. Others give no source.

For definiteness let me write down a proof, not in Cantor's words, which contains all the points we shall need to comment on.

(1) We claim first that for every map f from the set $\{1, 2, \dots\}$ of positive integers to the open unit interval $(0, 1)$ of the real numbers, there is some real number which is in $(0, 1)$ but not in the image of f.

(2) Assume that f is a map from the set of positive integers to $(0, 1)$.

(3) Write

$$0\,.\,a_{n1}\;a_{n2}\;a_{n3}\;\dots$$

for the decimal expansion of $f(n)$, where each a_{ni} is a numeral between 0 and 9. (Where it applies, we choose the expansion which is eventually 0, not that which is eventually 9.)

(4) For each positive integer n, let b_n be 5 if $a_{nn} \neq 5$, and 4 otherwise.

(5) Let b be the real number whose decimal expansion is

$$0\,.\,b_1\;b_2\;b_3\;\dots\,.$$

(6) Then b is in $(0, 1)$.

(7) If n is any positive integer, then $b_n \neq a_{nn}$, and so $b \neq f(n)$. Thus b is not in the image of f.

(8) This proves the claim in (1).

(9) We deduce that there is no surjective map from the set of positive integers to the set $(0, 1)$.

(10) Since one can write down a bijection between $(0, 1)$ and the set of real numbers (and a bijection between the positive integers and the natural numbers, if we want the latter to include 0), it follows that there is no surjective map from the set of natural numbers to the set of real numbers.

(11) So there is no bijection between these two sets; in other words, they have different cardinalities.

This is the proof which all the authors attacked. Most authors had seen a form of the argument which uses a picture: we write out the decimals $f(1)$, $f(2), \dots$ in a column with $f(1)$ at the top, and we trace out the real number b as we walk down the diagonal line

$$a_{11}, a_{22}, \dots\,,$$

changing the digits as we go. I shall call this the *written list* form of the argument.

None of the authors showed any knowledge of Cantor's theorem about the cardinalities of power sets.

§3. Why this target? Cantor's argument is short and lucid. It has been around now for over a hundred years. Probably every professional mathematician alive today has studied it and found no fallacy in it. So there is every temptation to imagine that anybody who writes a paper attacking it must be of dangerously unsound mind. One should resist this temptation; the facts don't support it. On a few occasions I was able to speak to the authors of these papers; one or two were clearly at sea, but others were as sane as you or me. In the course of researching this paper I came across statements by two of the leading logicians of this century, which—read literally—were just as crazy as anything in these attacks on Cantor's argument. Read on and judge.

There is a point of culture here. Several of the authors said that they had trained as philosophers, and I suspect that in fact most of them had. In English-speaking philosophy (and much European philosophy too) you are taught not to take anything on trust, particularly if it seems obvious and undeniable. You are also taught to criticise anything said by earlier philosophers. Mathematics is not like that; one has to accept some facts as given and not up for argument. Nobody should be surprised when philosophers who move into another area take their habits with them. (In the days when I taught philosophy, I remember one student who was told he had failed his course badly. He duly produced a reasoned argument to prove that he hadn't.)

To anticipate for a moment, I don't think any of our authors located anything distinctively bad about Cantor's argument. The points on which they tripped up were all things that might have tripped them in a thousand other more mundane arguments. Most of the muddles were not even mathematical. Different authors made different attacks.

It's nothing more than a guess, but I do guess that the problem with Cantor's argument is as follows. This argument is often the first mathematical argument that people meet in which the conclusion bears no relation to anything in their practical experience or their visual imagination. Compare it with two other simple facts of cardinal arithmetic. First, $m \times n = n \times m$. We can see what this amounts to by thinking of a rectangle with one side of length m and one side of length n. The picture points to the right formal argument when m and n are finite, and exactly the same argument works when they are infinite. Or second, $1 + \omega = \omega$. We don't meet ω in our everyday life, but we can see how to prove the inequality by moving each number along by one. (The picture lies well within the range of what we can 'übersehen', to quote Gödel [14].)

But then we come to Cantor's result, and all intuition fails us. Until Cantor first proved his theorem ([6], by a much longer argument, as it happens), nothing like its conclusion was in anybody's mind's eye. And even now we accept it because it is proved, not for any other reason.

§4. Not attacking an argument. It was surprising how many of our authors failed to realise that *to attack an argument, you must find something wrong in it*. Several authors believed that you can avoid a proof by simply doing something else.

The commonest manifestation was to claim that Cantor had chosen the *wrong enumeration of the positive integers*. His argument only works because the positive integers are listed in such a way that each integer has just finitely many predecessors. If he had re-ordered them so that some of them come after infinitely many others, then he would have been able to use these late comers to enumerate some more reals, for example the real number b which we defined in (5) of the proof.

Other authors, less coherently, suggested that Cantor had used the *wrong positive integers*. He should have allowed integers which have infinite decimal expansions to the left, like the p-adic integers. To these people I usually sent the comment that they were quite right, the set of real numbers does have the same cardinality as the set of natural numbers in *their* sense of natural numbers; but the phrase 'natural number' already has a meaning, and that meaning is not theirs.

One or two authors were ready with a counterargument. To say that the existing concept of natural numbers is incompatible with their numbers is to say that at least one of their numbers can't be included in the set of natural numbers. But we can demonstrate that any object whatever can be included in a natural number series. (Read Benacerraf [4], these authors might have added—though they didn't.)

This already goes to quite a deep issue about the identity of mathematical structures. I think there might be some difficulty in putting together an answer which everybody working in the foundations of mathematics would accept. But really the question should never have arisen in this context. There is no way that one can regard Cantor's assumptions about natural numbers as a mistake in his argument. The existence of a different argument that fails to reach Cantor's conclusion tells us nothing about *Cantor's* argument.

How does anybody get into a state of mind where they persuade themselves that you can criticise an argument by suggesting a different argument which doesn't reach the same conclusion?

Well, roughly as follows. Suppose our friend Hugo offers us a proof, by induction on n, that for every natural number n a man with n hairs on his head is bald. There are three degrees of response. The most passive is to say 'There must be a mistake somewhere', and leave it to somebody else to find where the mistake is. (In practice 'passive' is perhaps the wrong word, if we need to do some work to wall off a safe area of arguments where we never have to consider Hugo's.) The next is to look at Hugo's argument and try to find a place where Hugo has made a step which is not cogent. The third response, the most masterful, is to claim that one step in Hugo's argument is wrong *simply because the proof won't work without it*.

We have all seen responses of these three kinds to the paradoxes. Barwise and Etchemendy are forceful advocates of the second response in the introduction to their book on the liar paradox ([3] p. 7):

> A treatment of the Liar all too often takes the following form. First, various intuitively plausible principles are set out and motivated by a discussion of the commonsense notions involved. Then a contradiction is shown to follow from these intuitive principles. At this point the discussion turns directly to the question of which principles can be kept and which must be abandoned ... the Liar has forced us to abandon intuitively plausible semantic principles without giving us a reason, beyond the paradox itself, to suspect their falsehood. We see *that* they are false, without understanding *why*.

Readers of the proof of Theorem 6 on their page 79 can judge whether they might also be an illustration of the third response. (They motivate various intuitively plausible principles, short of one which they reject by an argument that 'exactly parallels the reasoning usually taken to show that the Liar is paradoxical'.)

I see no difference of principle between what these critics of Cantor are doing and what I called the masterful response above. We dislike the conclusion, so we outlaw one of the steps that got us there. Some might feel that on the moral scale there is a difference between a conclusion which is downright paradoxical and one that we happen to dislike. But I can't see how the response is justified in one case if it isn't in the other.

§5. Attacking an argument. In formal logic we teach people how to construct arguments, and how to check the validity of a formal argument. But we hardly teach anything about how to assess the cogency of an unformalised deductive argument. Our authors are making their criticisms without the benefit of any training in how to do it.

There are some good books about how to assess unformalised arguments. I have by my hand Alec Fisher's excellent volume 'The logic of real arguments' [11], and I shall quote it later. Like most of the genre, Fisher concentrates on scientific and moral arguments rather than deductive ones. Thus (p. 140):

> ... it is clearly difficult to *apply* [traditional formal logic] to *real* arguments—to arguments of the kind one finds for example in newspapers, magazines and learned journals.

If Fisher means to imply that traditional formal logic, as traditionally presented, gives the right tools for analysing informal *deductive* arguments, then part of my purpose is to sow some doubts about this.

This may be the moment to mention a passage in Wittgenstein's book 'Remarks on the foundations of mathematics' [22], where he claims (if I follow him right) that Cantor's argument has no deductive content at all. The theme of Wittgenstein's book is that mathematical statements get any meaning they may have from rule-governed activities that involve them. He singles out Cantor's argument because it would appear to have no relation to any imaginable activity.

Except for one, namely the activity of writing out lists of complete decimal expansions of real numbers. This is of course a daft activity, doomed to failure. Ah, says Wittgenstein, that's what Cantor's theorem must amount to ([22] p. 129):

> Surely—if anyone tried day-in day-out 'to put all irrational numbers into a series' we could say: "Leave it alone; it means nothing; don't you see, if you established a series, I should come along with the diagonal series!" This might get him to abandon his undertaking. Well, that would be useful. And it strikes me as if this were the whole and proper purpose of this method. It makes use of the vague notion of this man who goes on, as it were idiotically, with his work, and it brings him to a stop by means of a picture.

None of our authors showed any knowledge of Wittgenstein's critique, or any sympathy with it. They all regarded Cantor's argument as an attempt at a deductive proof of a meaningful proposition, and they all assessed it in these terms.

So how does one assess an unformalised deductive argument? Broadly speaking, such an argument has three kinds of component:

• There are the stated conclusion, the stated or implied starting assumptions, and the intermediate propositions used in getting from the assumptions to the conclusion. I shall call these the *object sentences*.

• There are stated or implied justifications for putting the object sentences in the places where they appear. For example if the argument says 'A, therefore B', the arguer is claiming that B follows from A.

• There are instructions to do certain things which are needed for the proof. Thus 'Suppose C', 'Draw the following picture, and consider the circles D and E', 'Define F as follows'.

A criticism of an argument might focus on any of these components. For example it might claim that one of the object sentences is meaningless or ambiguous; this would be an attack on the object sentences. It might claim that an object sentence appears somewhere without proper justification; this would be an attack on the justifications. It might claim that one of the things we are instructed to do in the proof is impossible; this would be an attack on the instructions.

In fact none of the authors took issue with the object sentences themselves, but there were several attacks on the justifications and the instructions.

One author did find another form of attack which I must mention. Sad to say, it was a flash of unintended brilliance, buried beneath a dozen pages in which nothing happened. This author had in front of him a form of Cantor's argument which used reductio ad absurdum. (In some formal systems this would be needed to pass from (8) to (9) in the proof above.) Let us prove, he said, that Cantor's argument is invalid. We start by assuming that it is valid. If it is valid we are entitled to use it; and so we do, down to the point where we get a contradiction. But since we have reached a contradiction, our original assumption must have been wrong. That is to say, Cantor's argument is invalid.

There is a quick though slightly dishonest refutation of this critique. Namely, Cantor's proof also makes an assumption, and when our author reaches the contradiction he only knows that at least one of his assumptions must be false; it need not be the one he made first. This refutation is dishonest, because it fails to point out that the assumption 'Cantor's proof is valid' doesn't play any role in the argument which follows. We are in territory quite close to Carroll's 'What the tortoise said to Achilles' [8], and I leave it to the reader to sort out the details. Typically, this refutation of Cantor's argument has nothing to do with Cantor's argument in particular—if it worked at all, it would work against any argument by contradiction, including those which the intuitionists find valid.

§**6. Attacks on the justifications.** In a well-respected textbook recently I noticed this sentence:

> In all cases $(0, 0)$ is a point of order 2 since any point of order 2 has the form $(x, 0)$, where x is a root of the cubic equation $0 = x^3 + ax$.

This looks very like an attempt to argue '$P(a)$, because for all x, if $P(x)$ then $Q(x)$'. It's extravagant to suppose that the author made a mistake of logic. More likely he meant to say 'since the points of order 2 are exactly those of the form … ', and he wasn't too careful about the exact wording because he expected the readers to think it through in their own terms anyway. This is a kind of conversational looseness. I doubt if Alfred Tarski was ever guilty of it, but probably most of the rest of us have been on occasion.

It's quite a different matter where a writer directly addresses the question whether Q follows logically from P, and gets it wrong. There were two of our authors who said they disagreed with Cantor about what follows directly from what.

The first of these authors denied the step from (8) to (9) in the proof. In fact he agreed that Cantor had proved that

> The image of any map from the set of positive integers to the set $(0, 1)$ is a proper subset of $(0, 1)$.

But he denied, several times over, that it follows that

> There is no injective map from the set of positive integers to the set $(0, 1)$, whose image includes all of $(0, 1)$.

(I repeat the caution that I'm not using the author's exact words. But he was reasonably proficient in set theory, and he should accept these quoted statements as equivalent to his formulations.) On the face of it, this author is denying that the inference

$$(*) \qquad \forall x\, \exists y\, \neg\phi(x, y) \vdash \neg\exists x\, \forall y\, \phi(x, y)$$

is valid.

The second author maintained that Cantor had proved something so strong that the result was paradoxical, though Cantor had failed to recognise this. He claimed that (8) directly implies

> The number b is not in the image of any map from the positive integers to $(0, 1)$.

He had no trouble in showing that this is absurd. I suppose he was using the fallacious inference

$$(**) \qquad \forall x\, \exists y\, \phi(x, y) \vdash \exists y\, \forall x\, \phi(x, y).$$

This fallacy is familiar from examples of the form 'Everything has a cause; therefore there is something that causes everything'.

There don't seem to be any recognised systems of logic in which $(*)$ is invalid or $(**)$ is valid. So I suppose these are just mistakes, not evidence for variant logics. I looked to see whether the psychological literature on mistakes of logic could throw any light. First let me quote Lance Rips' ([19] p. 392) list of factors which cause errors in experimental tests of logical reasoning:

> If Q follows from P according to some logical theory T but subjects fail to affirm that Q follows from P, that could be because (a) T isn't the appropriate normative standard; (b) subjects interpret the natural-language sentences that are supposed to translate P and Q in some other way; (c) performance factors (e.g., memory or time limits) interfere with subjects' drawing the correct conclusion; (d) the instructions fail to convey to subjects that they should make their responses on the basis of the entailment or deducibility relation rather than on some other basis (e.g., the plausibility or assertibility of the conclusion); (e) response bias [i.e., subjects' guesses about how the experimenter set up the test] overwhelms the correct answer; or (f) the inference is suppressed by pragmatic factors (e.g., conversational implicatures). If Q does not follow from P according to T but subjects affirm that Q follows from P, that could be because (a)–(e) hold as above; (g) subjects are interpreting the task as one in which they should affirm the argument, provided only that P suggests Q, or P makes Q more likely,

> or P is inductive grounds for Q; (h) subjects treat the arguments as an enthymeme that can be filled out by relevant world knowledge; (i) subjects ascribe their inability to draw the inference to performance factors and incorrectly guess that P entails Q; or (j) subjects are misled by a superficial similarity to some valid inference from P' to Q' into supposing that there is a valid inference from P to Q.

Is this list helpful? We have already ruled out the possibility that the authors were calling on some variant logic, or that they were assessing the argument as anything but a strictly deductive one. This disposes of (a), (d), (f), (g) and (h). The authors weren't up against limits of time, and they didn't regard Cantor's argument as an experimental test of their reasoning powers; so out go (c), (e) and (i). This leaves (b) and (j).

Case (b) would apply if the authors misinterpreted some sentence in Cantor's argument. Looking at what they say, I am sure this is not what has happened. In fact their mistakes are about the logical relations between sentences which they themselves have written.

Case (j) is obscurely stated. Does Rips mean that the subjects have misread the inference as being of some other form which happens to be similar? Or does he mean that they have correctly identified the form, but incorrectly guessed that the form must be valid because a similar one is? Either way round, I don't see why Rips gives this only as a cause of mistakenly inferring, not as a cause of mistakenly failing to infer.

What is missing from Rips' first list (a)–(f) is the case where a person correctly understands two sentences but fails to notice the logical connection between them.

Some writers have argued that if B follows logically from A and a person really understands both A and B, then that person *must* see that B follows from A. (For example, one could make a case that this is a criterion of whether the person 'really understands' the two sentences.) This was never plausible for the cases where it takes a lengthy argument to get from A to B. If we can fail to notice distant logical relationships, then it must at least be possible for us to fail to notice close ones. (And of course it happens. A few years ago an algebraist, now dead, published a long paper which seemed to be a major contribution to an important problem. His argument depended on finding a family of numbers which satisfy a certain very large set of equations and inequalities. Sadly it came to light that a particular small subset was unsatisfiable; the fact was obvious once it had been pointed out, but it was easily missed.) Though this is speculation, it seems to me the most natural explanation of our author's failure to accept the entailment $(*)$.

The corresponding explanation of $(**)$ would be that the author failed to notice the difference between two of his formulations; he thought he was paraphrasing when in fact he was reversing two quantifiers. I have an impression that the author may have been thrown by the fact that the diagonal number is always called b (in our proof), as if it was independent of f. One symbol, one object.

Rips' book contains a wealth of information about the errors that people do make in logical reasoning. I had hoped that he would give some data on $(*)$ and $(**)$. Unfortunately he chooses to 'represent' sentences by first finding equivalent sentences in prenex form and then using Skolem functions for existential quantifiers ([19] pp. 90ff, 185ff). Thus $(*)$ and $(**)$ become respectively

for $(*)$ $$\forall x \neg\phi(x, a_x) \vdash \forall x \neg\phi(x, a_x)$$

and

for $(**)$ $$\forall x\ \phi(x, a_x) \vdash \forall x\ \phi(x, a).$$

The steps involved in these reductions are at least as elaborate as either $(*)$ or $(**)$ on their own. The reduction of $(*)$ removes everything of interest. Rips doesn't mention any experimental tests of the reduced form of $(**)$; probably it's too bland to be tested. So Rips' book left me disappointed.

I turned next to the rival work of Johnson-Laird and Byrne [16]. This book I read with caution. I know I am not alone in finding its accounts of logical theory almost incomprehensible (see my brief review in [15]). Nevertheless the book does report some very interesting experiments.

Johnson-Laird claims in [16] and elsewhere that our normal mode of deductive reasoning is proof by cases; that we represent the cases by what he calls 'models' (they are not what model theorists call 'models'); and that we have no systematic procedure for finding the needed cases. A major cause of mistakes in deduction is failure to find the right cases. The more cases are needed, the more mistakes people make.

Johnson-Laird and Byrne [16] have a chapter (Chapter 7) on 'Many quantifiers: reasoning with multiple quantification'. This should be the place to find some treatment of $(*)$ and $(**)$; in particular we look there to find what Johnson-Laird and Byrne think the relevant 'models' are. But it transpires that all the examples in that chapter have the form

$$Q_1 x\ Q_2 y\ x\ R\ y$$

where Q_1, Q_2 are relativised quantifiers (like 'all of the musicians', 'some of the authors') and R is known to be an equivalence relation. We get the 'models' by sketching some equivalence classes and putting markers to show (a) what types of person or object occur in each and (b) which of these types are universally quantified. The result is a kind of Venn diagram with quantifiers. I didn't see how to extend this format to our situation.

A later chapter in the book (Chapter 9) claims to describe a procedure for constructing 'quantified models'. Much is obscure; the authors limit themselves to a language in which the only relation symbol is equality. But one can imagine that a refinement of their procedure would throw up the standard four-element counterexample to $(**)$ as a model of the sentence on the left. I suppose that the Johnson-Laird position would be that the author who thought $(**)$ was valid had generated only 'models' of the premise which were also 'models' of its conclusion, and failed to realise that other cases are possible. The problem (and I think it is Johnson-Laird's problem, not ours) is to bring this claim to a form which is (i) testable and (ii) significantly different from the bald statement that the author or the experimental subject has failed to realise that the conclusion doesn't follow from the premise.

Turning to the valid inference $(*)$, we run into a new problem. Johnson-Laird can explain how people make false inferences by failing to consider all the cases. But it was not at all clear to me how his theory explains people's failure to make correct inferences, or how he reaches any predictions about how hard people will find it to perform one or another correct deduction. Johnson-Laird and Byrne do discuss in detail an example of 'suppression of valid deductions' ([16] p. 81ff). But this turns out to be an example where a pragmatically misleading second premise causes the subjects to misinterpret the first premise, a phenomenon which seems to have nothing to do with the analysis in terms of 'models'. In sum, Johnson-Laird and Byrne also left me disappointed.

§7. **Attacks on the instructions.** This brings us to the third point of attack, the instructions.

One author complained that Cantor's proof requires us to write out an infinite diagram. But that's a thing we can't do; the author conscientiously proves this as follows. As we make the list, it becomes infinite either gradually, or suddenly, or not at all. The idea that it becomes infinite gradually is incoherent; at any stage it is either definitely finite or definitely infinite. If it suddenly becomes infinite, there is a stage at which it becomes infinite. But this is false; at every stage in the construction of the list, it is finite. Therefore it never becomes infinite.

Of course nobody would suggest that in order to carry out Cantor's proof you actually have to *write out* the infinite diagram, would they? Would they?

> Now suppose that
>
> $$x_0,\ x_1,\ x_2,\ x_3,\ \ldots$$
>
> is an infinite list or enumeration of some but not necessarily all of the real numbers belonging to the interval. Write down one below another their respective non-terminating decimal fractions ... [and here follows a diagram with some dot-dot-dots].

This is from Kleene's 'Introduction to metamathematics' ([18] p. 6). Taken literally, what Kleene says is quite mad. Of course one can avoid taking it literally by saying something like (3) in my version above. But it was clear that many of the authors had difficulties passing from the written list version of the argument to something more abstract. With hindsight it may have been unkind of Kleene to dump this on the unsuspecting beginner, six pages from the start of his book.

We move on quickly. A common fault in arguments is to take for granted something which should have been proved. One of our authors accused Cantor of doing this; he complained that Cantor *assumes that there is a map from the positive integers to* $(0, 1)$. See (2) in our version above.

It should be easy to deal with this. Cantor is assuming P in order to prove something of the form 'If P then Q'. This is a standard move in arguments. One assumes P 'for the sake of argument'. Nobody interprets this kind of assumption as a claim to *know* P, or even to *believe* P, do they? Do they?

Evert Beth is one of the few logicians who have seen problems in this form of argument and taken them seriously. He reports his conclusions on pages 36f of [5]. On page 17 of the same essay he had given a natural deduction argument where a premise numbered (2), viz.,

$$(Ey)[S(y)\&M(y)],$$

is assumed and then discharged later. Referring back to that argument he comments

> ... the (possibly false) assumption, which at a certain moment has been introduced, is eliminated later on ... However, if we wish exactly to know what is going on, then we ought to consult the semantic tableau. In the formal derivation of Section 4, we know by premiss (2), that some individual fulfils the condition $S(y)\&M(y)$, and we agree to give this individual the name 'a'.

This last sentence is completely mad. Beth implies that we know that premise (2) is true. But in the first place, nowhere in the article does he give any evidence whatever that (2) is true; in fact he has described it as 'possibly false' just a few lines earlier. In the second place, it is a string of symbols in an uninterpreted language (as far as we know—Beth has explained on p. 11 how to interpret a language, but he has said nothing to suggest that he has in mind any particular interpretation for this one); so no question of truth or falsehood arises. The last clause of the sentence is mad too: if we know that 'Some individual lives in Neasden', it makes no sense to "agree to give this individual the name 'a' " until we have picked out one such individual. But Beth has done nothing to pick out an individual.

Beth's mistakes here seem to me of the same order as any made by our critics of Cantor. He gets away with it because he is a brilliant logician, he writes a convincing style and we believe his conclusions. Though I have him in my sights here, probably most of us have said or written equally crazy things at one time or another. And in this particular case, Beth's account has the merit of highlighting two of the main problems about assumptions 'for the sake of argument'. First, when we assume P, we proceed *as if* we knew P. Second, when we assume there is x such that Px, we proceed *as if* we have identified such an x. What is going on here?

This is not the place to answer that question at length. But let me put down some pointers in this strangely uncharted territory.

Assumptions in arguments appear in at least the following four guises:

(a) The writer says 'I assume P because we already know P'. Here the assumption serves as a lemma.

(b) The writer says 'In the following diagram, assume A is the such-and-such, B is the such-and-such' etc., and then uses the diagram.

(c) The writer says 'Assume P', deduces Q, and concludes 'If P then Q'.

(d) The writer says 'Assume P', deduces something known to be false, and concludes 'Not P'. (Or the 'nots' could be the other way round.) This is reductio ad absurdum.

This list is not complete. For example I am ignoring the ancient and renaissance rule of false position ('regula falsi'), where we solve an equation by making two possibly incorrect guesses about the solution and then calculating the errors; see Smith [21] p. 437ff.

Form (a) is unproblematic and I say no more about it.

Form (b) occurs most often in geometric arguments, but one meets it elsewhere. On the Johnson-Laird theory we use a version of it all the time. Gelernter [13] makes it the basis for a computer implementation of geometric reasoning. People have found it problematic from earliest times, because the objects in the diagram might have different properties from the things that they represent. (Maybe we are proving properties of equilateral triangles, but our hand slips and the triangle we draw is scalene.) Aristotle raises the matter briefly in his Metaphysics [1] (book XIV 1089a24ff); his view seems to be that there is no harm done as long as the false assumption is not 'in' the proof. No doubt one can elaborate this into a reasonable theory, though there may be more to say in particular cases.

Natural deduction conflates the two forms (c) and (d). Leaving aside the cases which worry the intuitionists, it's agreed that both forms of argument are valid—these rules won't let us down. The problem is to explain (c) and (d), not as formal rules but as meaningful pieces of discourse. Until we can do this, I'm not sure that we have given a just and fair answer to the author who criticised step (2) of Cantor's argument.

There is a chapter entitled 'Suppose for the sake of argument that ...' in Fisher's book already mentioned [11] (Chapter 6). Fisher gives many sensible examples, and maybe they would be enough to soften the heart of Cantor's critic. Thus he comments:

> A mathematician who presents the standard Euclidean proof that there are infinitely many prime numbers begins by supposing that

> *there are only finitely many*. He is not asserting (telling us) that there *are* only finitely many primes (because he knows full well that this is false) but he is asking us to consider the proposition with a view to drawing out its implications.

Nevertheless I think there is something missing. Fisher has told us the *purpose* of forms (c) and (d). But there are other ways of achieving this purpose. We can draw out the implications of a proposition P without assuming P. For example we can use Frege's preferred style and stick to the format 'If P then . . . '. What Fisher has not told us is, first, exactly what we are being told to do when the argument says 'Assume . . . ', and second, why this is a good way of achieving the stated purpose.

Writers of a psychological cast sometimes speak of assuming 'for the sake of argument' as a kind of mental activity. Thus Rips ([19] p. 7f) conflates it with 'imagining a situation'. Some form of this view must be correct. For example in a debate one speaker may say to the other:

(+) When you say 'Q and R', are you assuming that P?

Normally the second speaker understands the question and knows whether the correct answer is Yes or No. 'Assuming' is something that we do with our minds, and normally we can tell whether we are doing it.

But this activity of assuming has some odd properties. First of all, we can assume things that we could never conceivably imagine. Thus to prove that there is no greatest integer, we start by assuming that there is one. If any reader knows how to imagine that there is a greatest integer, I'd be interested to hear how they do it and what it feels like. But in any case, this approach to assuming must be barking up the wrong tree. The validity of an argument can never depend on you or me doing some particular thing in the privacy of our imaginations. (Our imaginations might help us to *find* a valid argument, but this is a different matter.)

Second, in the debate just mentioned, the second speaker could quite meaningfully answer the question (+) by saying

> I am assuming it for Q but not for R.

So assuming is not a thing that we do at a particular time; it's a thing that we do at a particular stage in an argument, and in respect of certain things in the argument. This pulls 'assuming' out of the world of brute facts, and gives it an intentional or juridical feel.

Third, one can assume things which are not even meaningful propositions, since they contain symbols for which no reference has been given. The extreme case of this is where one makes assumptions in natural deduction arguments, using an uninterpreted first-order language, as in the example from Beth above. But there is already an example at (2) in our proof of Cantor's theorem, where we assume that f has some property without taking any steps to specify what f is.

One possible response is that the letter 'f' is a variable bound by an implied universal quantifier 'For all f'. The problem with this approach is that the scope of the quantifier would have to reach all the way down to clause (7) of the proof, through several sentences, including both statements and definitions. To make sense of this, we would need a semantics for discourse rather than for sentences one at a time. The semantics of discourse is still in its infancy. (See Kamp and Reyle [17] for a pioneering attempt.)

Besides the questions what (a)–(d) are separately, we can also ask how they are related. Beth (*loc. cit.*) takes for granted that (b)–(d) are all examples of the same phenomenon. A revealing passage in Michael Dummett's book on Frege's philosophy of language suggests links between (b)–(d) and the activity of *assigning a reference*:

> If, for example, I take some colour counters, and say, 'Let this one stand for the Government, this one for the Opposition, this one for the Church, this one for the Universities, this one for the Army, this one for the Trade Unions, ... ', and so on, I shall be understood on the presumption that I am about to make some arrangement of the counters by means of which I intend to represent some relations between these institutions, and assert that they obtain. If I do not go on to make any such arrangement, but simply start talking about something else, my earlier declarations lose their original intelligibility ... it is like my saying, 'Suppose there is life on Mars', and then failing to draw any consequences from this hypothesis, and, when challenged, saying, 'Oh, I simply wanted you to suppose that'; ...

([10] p. 193).

I think I know broadly what are the right answers to these questions, but I don't propose to argue the matter here. The conclusion I want to leave on the table is that the notion of assumptions in arguments is surrounded with serious philosophical puzzles. It can trip up a professional almost as easily as a beginner.

§8. Conclusion. First, contrary to what several critics of Cantor's argument suggested in their papers, at least one mathematician was prepared to look at their refutations with some care and sympathy.

Second, a small number of the criticisms are fair comment on misleading expositions. A much larger number of the criticisms are fair comment on some serious and fundamental gaps in the logic that we teach. Even at a very elementary level—I'm tempted to say *especially* at a very elementary level—there are still many points of controversy and many things that we regularly get wrong.

Third, there is nothing wrong with Cantor's argument.

REFERENCES

[1] ARISTOTLE, ***Metaphysics* X–XIV**, Loeb Classical Library, Harvard University Press, Cambridge, Mass., 1935.

[2] JOHN D. BARROW, ***Theories of everything: the quest for ultimate explanation***, Oxford University Press, Oxford, 1990.

[3] JON BARWISE and JOHN ETCHEMENDY, ***The liar***, Oxford University Press, New York, 1987.

[4] PAUL BENACERRAF, *What numbers could not be*, ***Philosophy of mathematics, Selected readings*** (Paul Benacerraf and Hilary Putnam, editors), Cambridge University Press, Cambridge, 2nd ed., 1983, pp. 272–294.

[5] E. W. BETH, *Semantic entailment and formal derivability*, ***The philosophy of mathematics*** (Jaakko Hintikka, editor), Oxford University Press, 1969, pp. 9–41.

[6] GEORG CANTOR, *Über eine Eigenschaft des Inbegriffes aller reellen algebraischen Zahlen*, ***Journal für die reine und angewandte Mathematik***, vol. 77 (1874), pp. 258–262.

[7] ———, *Über eine elementare Frage der Mannigfaltigkeitslehre*, ***Jahresbericht der Deutschen Mathematiker-Vereinigung***, vol. I (1891), pp. 75–78.

[8] LEWIS CARROLL, *What the tortoise said to Achilles*, ***Mind***, vol. 4 (1895), pp. 278–280.

[9] UNDERWOOD DUDLEY, ***Mathematical cranks***, Mathematical Association of America, Washington, DC, 1992.

[10] MICHAEL DUMMETT, ***Frege, philosophy of language***, Duckworth, London, 1973.

[11] ALEC FISHER, ***The logic of real arguments***, Cambridge University Press, Cambridge, 1988.

[12] ABRAHAM A. FRAENKEL, ***Abstract set theory***, 4th ed., North-Holland, Amsterdam, 1976, revised by Azriel Lévy.

[13] H. GELERNTER, *Realization of a geometry theorem-proving machine*, ***Proceedings of international conference on information processing***, UNESCO House, Paris, 1959, pp. 273–282.

[14] KURT GÖDEL, *Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes*, ***Collected works*** (S. Feferman et al., editors), vol. II, Oxford University Press, Oxford, 1990, pp. 240–251.

[15] WILFRID HODGES, *Critical commentary on P. Johnson-Laird and R. Byrne, 'Deduction'*, ***Behavioral and Brain Sciences***, vol. 16 (1993), p. 353f.

[16] P. N. JOHNSON-LAIRD and RUTH M. J. BYRNE, ***Deduction***, Lawrence Erlbaum, Hove, 1991.

[17] HANS KAMP and UWE REYLE, ***From discourse to logic***, Kluwer, Dordrecht, 1993.

[18] STEPHEN COLE KLEENE, ***Introduction to metamathematics***, North-Holland, Amsterdam, 1952.

[19] LANCE J. RIPS, ***The psychology of proof***, MIT Press, Cambridge, Mass., 1994.

[20] BERTRAND RUSSELL, ***The principles of mathematics***, George Allen and Unwin, London, 1903.

[21] D. E. SMITH, ***History of mathematics***, vol. II, Dover, New York, 1958.

[22] LUDWIG WITTGENSTEIN, ***Remarks on the foundations of mathematics***, Blackwell, Oxford, 1956.

SCHOOL OF MATHEMATICAL SCIENCES
QUEEN MARY AND WESTFIELD COLLEGE
MILE END ROAD
LONDON E1 4NS, ENGLAND, U.K.
E-mail: w.hodges@qmw.ac.uk