To appear in *Oxford Studies in Metaphysics*

On Infinite Size

Bruno Whittle

Late in the 19th century, Cantor introduced the notion of the 'power', or the 'cardinality', of an infinite set.[1] According to Cantor's definition, two infinite sets have the same cardinality if and only if there is a one-to-one correspondence between them. And what Cantor was able to show was that there are infinite sets that do not have the same cardinality in this sense. Further, since he equated the cardinality of a set with its *size*, he took this result to show that there are infinite sets of different sizes: and, indeed, this has become the absolutely standard understanding of the result. The aim of this paper, however, is to challenge this standard understanding—and, more generally, to argue that we do not, in fact, have any reason to think that there are infinite sets of different sizes.

I should underscore that I am not, in any way, going to challenge Cantor's *mathematics*: my arguments are aimed solely at the standard account of the *significance* of this mathematics. But I trust that the interest of the challenge is nevertheless clear: for, without this claim about significance, Cantor *cannot* be said to have established that there are different sizes of infinity.

The plan for the paper is as follows. In §1 I will give an initial argument against the claim that Cantor established that there are infinite sets of different sizes. This initial argument will proceed by way of an analogy between Cantor's mathematical result and Russell's paradox. Then, in §2, I will give a more direct argument against the claim that Cantor established that there are infinite sets of different sizes. Finally, in §3, I will consider objections to the arguments; and I will also consider what the consequences are, if they work.[2]

---

[1] See, e.g., Cantor [1883].

[2] I said that Cantor's equation of the size of a set with its cardinality has become absolutely standard. However, I should note that there *have* been challenges to this equation: in particular, there have been attempts to develop alternative accounts of infinite size on which two sets can be of different sizes even if there is a one-to-one correspondence between them; see, e.g., Mancosu [2009] and the work cited there. As

1. An Initial Argument

The aim of this section is thus to give an initial argument against the claim that Cantor established that there are infinite sets of different sizes.

I should start by stating (what I will refer to as) Cantor's account of infinite size. Thus, as I said in the introduction, according to Cantor's notion of cardinality, two infinite sets have the same cardinality iff there is a one-to-one correspondence between them. Further, Cantor equated the cardinality of a set with its size. Together, these claims thus yield the following account of when two infinite sets are of the same size.[3,4]

(C1) For any infinite sets A and B, A is the same size as B iff there is a one-to-one correspondence from A to B.

Further, according to Cantor's notion of cardinality, the cardinality of A is at least as great as the cardinality of B iff there is a one-to-one function from B to A.[5,6] Thus, once again equating claims about cardinality with claims about size (as Cantor did, and has become standard), we get the following.

---

will become clear, the challenge that I will raise in this paper is of a very different sort: it is a challenge to the claim that if two infinite sets are the same size, then there is a one-to-one correspondence between them, whereas the challenges just mentioned are to the converse of this claim. A thorough discussion of these alternative challenges is, unfortunately, beyond the scope of this paper. However, one reason why one might be *somewhat* sceptical about their prospects is that we do *seem* to be in possession of a very good argument for the claim that they challenge (i.e., the claim that the existence of a one-to-one correspondence between two sets entails that they are of the same size); see §2 below (although see Mancosu [2009] for a dissenting evaluation of a similar argument). In contrast, I will argue that the best arguments for the claim challenged *here* can in fact be shown to fail. However, I should also note that everything that I will say here could easily be made compatible with the success of these alternative challenges, if it turns out that they are successful.

[3] A function from A to B is a one-to-one correspondence iff: (i) any two members of A are sent to different members of B; and (ii) every member of B has some member of A sent to it.

[4] Cantor proposed not only (C1), but also its generalization to *all* sets (whether infinite or finite). For simplicity, I will initially focus only on the claim for infinite sets. But the claim for finite sets will be discussed further in §3.2.

[5] A function from B to A is one-to-one iff any two members of A are sent to different members of B.

[6] An alternative definition would say that the cardinality of A is at least as great as the cardinality of B iff there is an onto function from A to B; where a function from A to be B is onto iff every member of B has some member of A sent to it. These two definitions are equivalent, given the axiom of choice (which says that for any set C of disjoint sets, there is a set D that contains exactly one member of each member of C). For the purposes of this paper, I will assume that these two definitions are equivalent (but nothing that I will say will make essential use of this fact).

(C2) For any infinite sets A and B, A is at least as large as B iff there is a one-to-one function from B to A.

So by 'Cantor's account of infinite size' I will mean this pair of claims, (C1) and (C2).

Now, given this account of infinite size, to establish that there are infinite sets of different sizes, it suffices to establish that there are infinite sets A and B without a one-to-one correspondence between them. And this Cantor did with the following groundbreaking result.[7]

Cantor's Theorem. For any infinite set A, there is no one-to-one function from the powerset of A to A.

The proof of the theorem is then as follows.

Proof. Suppose that f is a one-to-one function from P(A) into A, and consider $C = \{x \in A: \exists y \in P(A)$ such that $f(y) = x$ and $x \notin y\}$. But now consider f(C). And suppose first that $f(C) \in C$. Then (by the definition of C, and the fact that f is one-to-one) it follows that $f(C) \notin C$. So $f(C) \notin C$. But then (by the definition of C again) $f(C) \in C$: which is a contradiction.

So that (allegedly!) is how Cantor established that there are different sizes of infinity. The aim of this section, however, is to give an initial argument against the claim that Cantor really established this. This initial argument is in terms of Russell's paradox, and the basic idea is as follows. There is a very close analogy between the proof of Cantor's theorem and the derivation of Russell's paradox: indeed, they are really just the same argument in slightly different settings. And, similarly, there is a very close analogy between the following two claims: (a) the claim that Cantor established that the powerset of A is always larger than A; and (b) the claim that the reason for—or the diagnosis of—Russell's paradox is that there are more pluralities than there are objects. Indeed, the analogy between Cantor's proof and the paradox is so tight that it would seem that these two claims must stand or fall together. However what I will give is an argument *against*

[7] If A is a set, then the powerset of A is the set of all of A's subsets. I will use P(A) for this set.

the claim about Russell's paradox: and this will thus give an initial argument against the claim about Cantor's result.

I will start, then, by giving the derivation of Russell's paradox. This is the proof of a contradiction from Frege's Basic Law V, which (in slightly updated form) is as follows; here (and throughout) uppercase 'X', 'Y', etc. range over pluralities, while lowercase variables range over objects; thus, in the law, 'ext' is a term intended to denote a function from pluralities to objects (and 'ext' stands for extension; so the idea is that ext(X) is the 'extension' of X).[8]

(V) $\forall X \forall Y(\text{ext}(X) = \text{ext}(Y) \leftrightarrow \forall z(Xz \leftrightarrow Yz))$

Thus, (V) says that ext is a one-to-one function from pluralities to objects (X and Y are the same plurality iff for any z, Xz iff Yz; so (V) says that ext(X) = ext(Y) iff X and Y are the same plurality). What the paradox shows, however, is that there can be no such function. For consider the plurality R, consisting of those objects x such that: for some plurality Y, ext(Y) = x and x is not in Y. First suppose ext(R) is in R: then (by the definition of R, together with the fact, from (V), that ext is one-to-one) ext(R) is not in R. So ext(R) is not in R. But then (by the definition of R again) ext(R) *is* in R: which is of course a contradiction. That, then, is the derivation of the paradox.

Clearly, the argument here is essentially just that of the proof of Cantor's theorem (with the plurality R defined here in just the same way that C was in that proof, and playing the same role in the argument). Thus suppose (in accordance with orthodoxy) that Cantor's argument does indeed establish that, for any infinite set A, A has more subsets than members. Then, presumably, what Russell's paradox shows—what the reason for the paradox is—is that there are, similarly, more pluralities than there are objects. And, assuming Cantor's account of infinite size, that is indeed a very natural diagnosis of the paradox. But—natural or not—we will see that it cannot be right: and we will see this by considering a variant of the paradox that is so similar to the original that it *must* have the same diagnosis; but, *also*, it will be completely clear that the diagnosis of the variant has

[8] Frege's original version of the law was about concepts rather than pluralities. I am stating it in terms of pluralities since these seem to raise fewer distracting issues. However, to remain relatively close to Frege's original version I will (inessentially) assume that there is an empty plurality (i.e., a plurality X such that for any z, $\neg Xz$).

nothing to do with size; and, in that case, it will follow that the diagnosis of the original *similarly* cannot have anything to do with size.

So the first thing is to give the variant paradox. Now, (V) attempts (in effect) to assign a distinct object to each plurality (i.e., each plurality is assigned an 'extension', and (V) says that distinct pluralities get distinct extensions). But now suppose that—inspired, perhaps, by the suggested diagnosis of Russell's paradox—we rein back our ambitions, and try instead merely to assign every *definable* plurality its own object. That is, suppose that all we try to do is to assign a distinct object to every plurality that is defined by a formula of our language. Thus, instead of (V), we propose the following; here φ(z) stands for a formula of our language (and so (V*) is a schema, with a different instance for each different formula).

(V*) $\forall X(\forall z(Xz \leftrightarrow \varphi(z)) \rightarrow \forall Y(\text{ext}(X) = \text{ext}(Y) \leftrightarrow \forall z(Xz \leftrightarrow Yz)))$

So (V*) is, in effect, the restriction of (V) to definable pluralities. And, if the problem with (V) was that there are more pluralities than objects, then presumably (V*) will be entirely unproblematic: because clearly there are not more *definable* pluralities than objects, because there are no more definable pluralities than there are formulas to do the defining; thus, since formulas just *are* objects, there are no more such definable pluralities than there are objects; and so (V*) should not be problematic in the way that (V) was.[9]

But unfortunately (V*) *is* problematic: and in just the same way that (V) is. That is, we can derive a paradox from (V*) in just the same way that we did from (V). For consider the plurality R, defined just as before: i.e., let R be the plurality of those objects x such that for some plurality Y, x = ext(Y) and x is not in Y. Then, as before, we have ext(R) is in R iff it is not: for suppose first that ext(R) is in R; then (by the definition of R, together with the fact that R is defined by the formula $\exists Y(z = \text{ext}(Y) \wedge \neg Yz)$, and the fact that ext is one-to-one for definable pluralities) we get that ext(R) is not in R; so ext(R) is

[9] In case one is unconvinced by the claim that formulas are objects, one could give a version of this argument using merely the fact that there will be a one-to-one correspondence between the formulas of our language and the natural numbers, together with the fact that if there is a one-to-one correspondence between two sets, then they are of the same size. (The latter is the direction of Cantor's account that I will *not* challenge; indeed, I will give an argument for this direction of the account in §2.)

not in R; but then (by the definition of R, again) we get that ext(R) *is* in R; so contradiction.

Thus, even if we restrict attention to definable pluralities, we still get a paradox. And, further, the paradox involves essentially the same argument that Russell's original paradox did. That is, the variant paradox is *extremely* similar to the original. And, given that, they are presumably going to have the same (or very similar) diagnoses. But—as we have seen—the diagnosis of the variant cannot have anything to do with size (because there are *not* too many definable pluralities to allow each to get its own object). And so it seems that the diagnosis of Russell's original paradox similarly cannot have anything to do with size. But (as we also saw above) the idea that Russell's paradox *should* be diagnosed in terms of size would seem to stand or fall with the claim that Cantor established that there are different sizes of infinity. So we seem to have an initial argument against the claim that Cantor established that there are different sizes of infinity.

That, then, concludes the work of this section. In the next section I will try to give a more direct argument against the claim that Cantor established that there are different sizes of infinity.

2. A More Direct Argument

So, the aim of this section is to give a direct argument to the effect that we are not justified in believing Cantor's account of infinite size (i.e., (C1) and (C2) of §1). Thus, the first question to ask is: what reason might we have for believing this account?

And one thought one might have here is the following.

> Surely (C1) simply states *what it is* for two infinite sets to be of the same size; and, similarly, surely (C2) simply states *what it is* for one infinite set to be at least as large as another. That is, surely the right-hand-sides of (C1) and (C2) simply unpack what it is for the relation mentioned in the left-hand-side to hold. So—similarly—surely we can justify our belief in (C1) and (C2) simply by reflecting on the nature of these relations mentioned in the left-hand-sides.

So perhaps there is a very simple and easy account of why we should believe Cantor's account? Unfortunately, though, tempting or not, this thought is hopeless. For the size of a set (infinite or otherwise) is an intrinsic property of that set: that is, it is a property that a set has purely in virtue of what *it* is like; it is *not* a property that it has in virtue of its relations to distinct sets, or to functions between it and such sets.[10] Thus, for two sets (infinite or otherwise) to be of the *same* size is simply for them to have a certain sort of intrinsic property in common; it is *not* for these sets to stand in some sort of relation to certain functions between the two sets. So (C1) does *not* state *what it is* for two sets to be of the same size. And, for analogous reasons, (C2) does not state what it is for one set to be at least as large as another.

Of course, this is not to say that (C1) and (C2) cannot still be true: but they cannot, it seems, be seen to be so simply by reflecting on the nature of the same-size relation, or on the nature of the at-least-as-large-as relation. So, I ask again: what reason might there be for believing Cantor's account of infinite size?

It is perhaps useful, at this point, to separate out the different directions of the claims of the account (and perhaps also to reproduce the claims).

(C1) For any infinite sets A and B, A is the same size as B iff there is a one-to-one correspondence from A to B.

(C2) For any infinite sets A and B, A is at least as large as B iff there is a one-to-one function from B to A.

In each case, the left-to-right direction goes from a claim about size to a claim asserting the existence of a function; while the right-to-left direction goes (of course) from the functional existence claim back to the claim about size. And, actually, we seem to have pretty good reasons for believing the latter directions of the account (i.e., the function-to-size claims). Consider, for example, the function-to-size direction of (C1): this can apparently be argued for as follows.[11]

> Suppose first that A is some infinite set, let x be some member of A, and let y be some object that does *not* belong to A. Now let A* be the result of removing x from A, and replacing it with y. Surely A* is the same size as A: for the *size* of a set does not depend

[10] To put this last point slightly more carefully: the size of a set depends only on which members a set has; it does not depend on its relations to sets *other than its members*, or on its relations to functions from it to other sets, etc.
[11] For a similar argument, see Gödel [1947: 176].

> on *which* members it has, just on *how many* it has; and so swapping one member for another should not affect size.
>
> But now suppose that B is some other set, and that there is a one-to-one correspondence from A to B. In this case, B is, in effect, the result of simultaneously replacing each member of A with a distinct object. And, just as in the A* case (and for similar reasons), it seems that B must thus be the same size as the set one started with; that is, it seems that A and B must be the same size here.

Thus, we seem to have good reason to believe the function-to-size direction of (C1); and a similar argument can be given for the corresponding direction of (C2) (but this time using the principle that a set is at least as large as each of its subsets).

But now what about the size-to-function directions of (C1) and (C2)? Ideally, at this point in the paper I would consider the best arguments that been given for these. Unfortunately, however, it is hard to find *any* arguments for these directions of the claims.[12]

Before one has thought much about infinite sets, one might be tempted to argue as follows. Suppose that A and B are infinite sets, and suppose (for the sake of argument) that they are the same size. Then surely one can just *construct* a one-to-one correspondence from A to B: simply first choose some member of A, then choose a member of B to send it to; then another member of A, and another member of B to send it to, and so on. Surely (the thought would go) if A and B are *really* the same size, then this will eventually yield a one-to-one correspondence: for if one runs out of members of A before one runs out of members of B (say), then surely that just shows that A is *smaller* than B (and similarly if one runs out of members of B first).

Unfortunately, though (although this argument looks fine in the finite case), it is of course hopeless in the infinite one. For, even if A and B are both the same set (for example, the set of natural numbers), there is no guarantee that simply choosing members will lead to a one-to-one correspondence (for example, suppose that one chooses members of 'A' in the obvious order, i.e., 0, 1, 2, etc., but that one chooses members of

[12] For one very striking failure to give such an argument, see (again) Gödel [1947]. Gödel starts the paper by asking if Cantor's account of infinite size is 'uniquely determined'. He then proceeds to give an argument for the function-to-size direction of this account (his argument is similar to that which I have just given). And he then concludes from this that Cantor's account *is* uniquely determined—without any apparent recognition of the fact that he has (in effect) just argued for a biconditional by arguing for one direction of it. Similarly, set theory textbooks typically present Cantor's account with very little in the way of accompanying argumentation (for example, Hrbacek and Jech [1999] simply mention a case involving theatregoers and seats, and then say that Cantor's account is 'very intuitive' ([1999: 65–66])).

'B' in the order 0, 2, 4, etc.; then one will obviously not end up with a one-to-one correspondence from A to B). So the idea that simply choosing members will lead to a one-to-one correspondence, as long as A and B are the same size, is hopeless.[13]

So that argument won't work. But it is not, I take it, that proponents of Cantor's account think that these size-to-function claims must simply be taken on faith. Rather, the thought (I take it) is that they are sufficiently obvious that we are entitled to believe them, even in the absence of any explicit argument (and I must confess that that is what I thought, when I first learnt set theory). But *why* should it seem obvious that if a pair of sets stand in a certain size-relation, then there should exist a certain sort of function between them—especially in light of the fact that it is far from obvious how to actually construct such a function? The implicit thought, surely, is something like the following (what I am about to say is about (C1), but it could easily be rephrased so as to be about (C2)):

> Let A and B be infinite sets, and suppose, for the sake of argument, that there is no one-to-one correspondence between them. Well, what possible reason could there be for *why* there is no such function? The only possible reason, surely, is that the two sets are of different *sizes*—for what else could be relevant here? That is, if there is no one-to-one correspondence between A and B, then they must be different sizes; but that is logically equivalent to the relevant direction of (C1) (i.e., that if A and B are the same size, then there *is* such a function).

Thus, I suggest that the reason *why* the size-to-function claims seem obvious is because of something like such an inference to the best explanation: i.e., the thought is that if there *isn't* a one-to-one correspondence between two sets, then the only possible (and hence the best!) explanation is that one set must be bigger than the other.[14]

What I will argue in the rest of this section, however, is that—natural as this thought may be—it is mistaken. For (I will argue) in the paradigm cases of pairs of infinite sets without a one-to-one correspondence between them, there *is* a better explanation for why there is no such function; so the upshot will be that we should no

[13] I will discuss an attempt to strengthen this argument, using the fact that every set has a least well-ordering, in §3.1 (but I will argue that the modified argument also fails).
[14] I will take for granted here that it makes sense to talk about explanations of mathematical facts (of course, if it doesn't make sense, then that would seem to make things even worse for Cantor). For a discussion of such explanations, see Mancosu [2011].

longer believe Cantor's account of infinite size (because the reason that it seemed obvious turns out to be a mistake).

In making my case, for the sake of definiteness I will focus on the set of natural numbers, N, and its powerset (but similar points could be made about any infinite set A and *its* powerset). Thus, what I will argue is that there is an explanation of why there is no one-to-one correspondence between N and P(N) that is better than that which uses the hypothesis that one is larger than the other.

So what is this better explanation? Well, the thought is simply as follows. Here is a completely banal and general fact of mathematical life: there are very often 'connecting principles' between mathematical domains, D_1 and D_2, which say that for every object d_1 of D_1, there is an object d_2 in D_2, that is related to d_1 in a certain way; and these principles are often pretty self-evident to anyone who understands the natures of the domains D_1 and D_2. So, to illustrate, here is an example where $D_1 = D_2 = P(N)$.

(1) For every set of numbers A, there is another set of numbers B, that contains precisely the numbers that are not in A.

This, surely, is pretty self-evident to anyone who knows what sets of numbers are. Another example where $D_1 = {}^{N}N$ (i.e., the set of functions from N to N), and $D_2 = P(N)$, is as follows.

(2) For any function f from N to N, there is a set of numbers A, that contains precisely those numbers that f sends to 0.

Again: surely pretty self-evident to anyone who knows what the domains D_1 and D_2 are. And now here is another example, where $D_1 = {}^{N}P(N)$ and $D_2 = P(N)$.

(3) For any function f from N to P(N), there is a set A that contains precisely those numbers that are not members of the sets that they are sent to by f.

Again, this principle is surely pretty self-evident to anyone who knows what functions from N to P(N) are, and also what sets of numbers are. But it turns out that this last principle gives us a completely sufficient explanation for why there is no one-to-one correspondence from N to P(N): for why, in particular, there is no onto function from N to P(N).[15] This explanation, starting with (3), is as follows. Let f be any function from N to P(N). Then, by (3), there is a set A containing precisely those numbers that are not members of the sets they are sent to (by f). And so suppose that for some n, f(n) = A; and suppose, to begin with, that n is in A. Then, by the definition of A, n is not in f(n), i.e., n is not in A: so it turns out that n is not in A, after all. But then, by the definition of A again, it follows that n is in A: which is a contradiction. So, given only the connecting principle (3), we can completely explain why there can be no onto function from N to P(N) (and thus why there cannot be a one-to-one correspondence from N to P(N)).

What I now want to argue is that this is in fact a *better* explanation than any in terms of the sizes of N and P(N). But I should first just make clear that this really is a *different* explanation. Actually, though, that is relatively obvious. For the explanation that I have proposed starts from a fact connecting functions from to N to P(N) (on the one hand) and members of P(N) (on the other); and this is simply a very different fact from the (alleged) fact that P(N) is bigger than N (for the latter, as we have seen, concerns only the intrinsic properties of N and P(N), and not functions between them); thus, the two explanations start from very different facts, and so they are different.

But—one might respond—

> OK. The explanation that you have proposed really *is* different. But only because it is incomplete: yes, it starts from something that is not (in and of itself) about size; but this starting point must itself be explained; and surely *that* explanation will bottom-out at a fact about the relative sizes of N and P(N).

This response is not very promising, however. For, while it may be correct that my explanation is incomplete (i.e., perhaps (3) must, as the respondent contends, itself ultimately be explained), nevertheless, it is hardly plausible that this explanation should take us back to the sizes of N and P(N). For, surely, whatever this ultimate explanation of (3) is going to look like, it is going to be essentially similar to the ultimate explanations of other connecting principles, such as (1) and (2). And surely one does not want to say

---

[15] The explanation I am about to give corresponds to the proof of an alternative version of Cantor's theorem (i.e., the version that says that there is no onto function from N to P(N)). The points that I will make could also be put in terms of an explanation corresponding to the proof of the version of the theorem in §2; however, it is slightly simpler to focus on the explanation that I do.

that every connecting principle from D_1 to D_2 must ultimately be explained in terms of the sizes of the domains D_1 and D_2 (or anything along those lines). Rather, it is surely much more plausible to say that these principles are explained by what sorts of things the members of the domains are (facts about the conditions for their existence, for example). For instance, in the case of each of (1–3), the ultimate explanations are plausibly all going to start from the fact that for any property of numbers, there is a set of numbers that contains precisely the numbers with that property.[16]

Thus, the proposed explanation of why there is no one-to-one correspondence from N to P(N) really does seem to be an alternative explanation to that in terms of size. But is it *better*? Well, *of course* it is: because it only uses things that we are committed to anyway (i.e., the connecting principle (3); for it is not as if, if we accept that P(N) is larger than N, then we would give up on (3)). That is, the proposed explanation is clearly more economical than that in terms of size. And, thus, since the only reason we could find for believing Cantor's account in the first place was a sort of inference to the best explanation, it would seem to follow that we do not, in the end, have any reason to believe that account.

3. Objections and Consequences

In this section I will consider objections my arguments; and I will also consider what the consequences are, if the arguments work.

3.1. Least Well-Orderings

Now, in §2 I considered an argument for Cantor's account based essentially on the following thought: surely if A and B are the same size then one can simply *construct* a one-to-one correspondence between them, i.e., by successively choosing members of the two sets. But I rejected this argument as hopeless: because even if A and B are the very

[16] This basic fact about sets of numbers may *also* explain *how many* such sets there are, but that in no way tells against what I am saying: for what is crucial is simply that this basic fact is not in and of itself a fact about size.

same set, there is no guarantee that the proposed construction will yield a one-to-one correspondence. One might, however, be tempted to respond as follows.

> OK, unlike in the finite case, it is not true that *any* way of choosing members will yield a one-to-one correspondence. But every set has a least well-ordering.[17] And as long as one chooses the members of A and B in line with least well-orderings of the sets,[18] then, as long as they really *are* of the same size, one will end up with a one-to-one correspondence. So Cantor was right after all!

Unfortunately, however, this response simply begs the question. For what this response is simply *taking for granted* is that if A and B are the same size, then their least well-orderings will be isomorphic. For if these well-orderings are *not* isomorphic then the construction described will not yield a one-to-one correspondence. But what is being taken for granted is then *stronger* that what we are trying to prove: because an isomorphism between $\prec_A$ and $\prec_B$ (where these are well-orderings of A and B, respectively) *is* (among other things) a one-to-one correspondence between A and B. Thus, the proposed argument simply begs the question.

3.2. What about the Finite Case?

An alternative objection focuses on the version of Cantor's account for finite sets. For (the objection goes) surely *this* version of Cantor's account is correct (i.e., surely for any *finite* sets A and B, they are the same size iff there is a one-to-one correspondence between them). But one might then worry that if my argument works in the infinite case, then it will work in the finite case too: giving the unacceptable result that we are not justified in believing even the finite version of Cantor's account. In fact, however, there is no need for concern here. For, as I have already hinted, it is actually clear how this

---

[17] A well-ordering of A is a relation $\prec$ on A that is anti-symmetric (if $x \prec y$ then not $y \prec x$), and such that every non-empty subset of A has a $\prec$-least element (i.e., for every subset X of A, there is $x \in X$ such that for every $y \in X$, if $y \neq x$ then $x \prec y$). The domain of a relation $\prec$ is the set of things x such that for some y, $x \prec y$ or $y \prec x$. And a relation $\prec$ is isomorphic to a relation $\prec'$ iff there is a one-to-one correspondence f between the domains of $\prec$ and $\prec'$ such that for any x and y in the domain of $\prec$, $x \prec y$ iff $f(x) \prec' f(y)$. Finally, $\prec$ is the least well-ordering of A iff for any well-ordering $\prec'$ of A, $\prec$ is isomorphic to an initial segment of $\prec'$.

[18] I.e., as long as for some least well-orderings $\prec_A$ and $\prec_B$ of A and B, respectively, one first chooses the $\prec_A$-least member of A, and sends it to the $\prec_B$-least member of B; and one then chooses the second $\prec_A$-least member of A, and sends it to the second $\prec_B$-least member of B; and so on.

version of Cantor's account can be supported. For, in the finite case, one *can* simply give the argument that is 'hopeless' in the infinite case: because given *finite* sets of the same size A and B, one *can* always construct a one-to-one correspondence between them simply by choosing successive members of them. Thus, in *this* case, there is no need to fall back on an inference to the best explanation, and, similarly, there is no danger that my challenge will generalize in the way that we were worried about.

3.3. Significance

A different sort of worry one might have about my arguments concerns not their cogency but their significance. For, one might think something like the following.

> OK, perhaps you're right that Cantor did not actually establish anything about *size*. Still, he did introduce a rich and fruitful mathematical concept (i.e., *cardinality*). And why should we really care if his results are really about *size*, as opposed to being merely about cardinality (which one might call 'size*')?

I must admit that I find this line of thought incredible: when I learnt (or came to believe!) that Cantor had shown that there are different sizes of infinity, I thought that it was one of the most exciting mathematical results I had ever encountered. I would not have been anywhere near as excited if all I had come to believe was that Cantor had provided a new technical notion (with certain similarities to the notion of size, perhaps) and shown that there are infinite sets which this new notion puts into different categories. Surely I am not alone in feeling this way!

Another way to put essentially the same point is this. Cantor's theorem surely belongs to a general category of mathematical results whose significance in large part depends on their connection to pre-theoretic notions. Another good of example of such a result is that of Kurt Gödel and Alonzo Church to the effect that the set of arithmetical truths is not computable (i.e., that no computer could output precisely the true sentences of the language of arithmetic).[19] Now, the way in which the Gödel-Church result is actually *proved* is by providing some precise mathematical definition of computability, and then showing that the set of arithmetical truths is not computable in this sense. But it

[19] For this result, see, e.g., Boolos and Jeffrey [1989: 176].

is surely obvious that the significance of the technical result depends in very large part on the adequacy of the definition of computability (i.e., on whether it is really coextensive with the pre-theoretic notion). And it is surely similarly obvious that the significance of Cantor's result depends in very large part on the adequacy of his technical notion of size (i.e., on whether or not what I have been calling his account of infinite size is correct). But, if that is right, then the significance of *my* conclusion should also be clear (because it gets to the heart of the significance of Cantor's fundamental result!).

3.4. Consequences

So much, then, for objections. I want to end the paper by saying something about what the consequences are, if the arguments that I have given are correct. And the first natural question to ask here is of course this: so are there, after all that, different sizes of infinity? For, if we accept that Cantor did not succeed in answering this question, then it is of course very natural to ask what the answer really is. Now, giving any sort of definite answer is well beyond the scope of this paper. However, I do want to suggest that—while we work on that!—the most reasonable view to take is that there is exactly *one* size of infinity. For, this is clearly the simplest hypothesis, and, if the above arguments work, then Cantor did not give us any reason to prefer an alternative. Of course, I am not suggesting that we are entitled to believe this one-size-hypothesis with anything like the certainty that previously we thought that we were entitled to believe Cantor's hypothesis. But the simplicity of the former does seem to give it a good claim to being our best working hypothesis.

I want to make one further point about the picture that emerges, if my arguments are correct. And I will for the sake of definiteness once again focus on N and P(N) (but similar points could be made about any infinite set and its powerset). Now, what Cantor *did* of course establish is that there is no one-to-one correspondence between N and P(N). But what I have been arguing is that this does not tell us anything about the sizes of the sets, because the reason *why* there is no such function is not that the sets are of different sizes; rather, the lack of such a function is due to basic connections between, on the one hand, functions between N and P(N), and, on the other, the members of P(N) (cf. (3) of

§2). That is to say: on the picture that emerges (if my arguments are correct), these functions cannot be used to *measure* N and P(N) because there is a basic connection between the functions and the sets that gets in the way. So, if I am right, then one way to think about the situation is this: these functions are simply not the 'independent observers' that Cantor needed them to be.

Thus, I hope to have shown that, for all we know, there is only one size of infinity.[20]

References

Boolos, G. and R. Jeffrey. 1989. *Computability and Logic*. Third edition. Cambridge: Cambridge University Press.

Cantor, G. 1883. Foundations of a General Theory of Manifolds: A Mathematico-Philosophical Investigation into the Theory of the Infinite. W. Ewald (trans.) in W. Ewald (ed.), *From Kant to Hilbert: A Source Book in the Foundations of Mathematics, Volume II*, 1996: 881–920. Oxford: Clarendon Press.

Gödel, K. 1947. What is Cantor's Continuum Problem? In his *Collected Works: Volume II, Publications 1938–1974*, S. Feferman *et al.* (eds), 1990: 176–87. Oxford: Oxford University Press.

Hrbacek, K. and T. Jech. 1999. *Introduction to Set Theory*. Third edition. New York: Marcel Dekker.

Mancosu. P. 2009. Measuring the Size of Infinite Collections of Natural Numbers: Was Cantor's Theory of the Infinite Inevitable? *Review of Symbolic Logic* 2: 612–46.

——2011. Explanation in Mathematics. In E. Zalta (ed), *The Stanford Encyclopedia of Philosophy*, Summer 2011 edition. URL = <http://plato.stanford.edu/archives/sum2011/entries/mathematics-explanation/>.

[20] For comments and discussion I am extremely grateful to George Bealer, Cian Dorr, John Hawthorne, Øystein Linnebo, Alex Paseau, Agustín Rayo, Zoltán Gendler Szabó and Tim Williamson, and also to audiences at Bristol, Oxford and Yale.