The Structure of Emptiness

Graham Priest

1 Introduction

The view that everything is empty (śūnya) is a central metaphysical plank of Mahāyāna Buddhism. It has often been the focus of objections. Perhaps the most important of these is to the effect that it entails a nihilism: nothing exists. The objection, in turn, is denied by Mahāyāna theorists, such as Nāgārjuna. One of the things that makes the debate hard is that the precise import of the view that everything is empty is unclear. The point of this paper is to revisit and resolve the debate. I will do this by giving a precise mathematical characterisation of emptiness, showing that emptiness, whatever it is, has a precise structure, and is not therefore, to be identified with the void. Thus the debate will be resolved in favour of the Mahāyāna.

2 Relational Existence

Before turning to the mathematisation, however, let us start with an informal exposition of the matters at hand. Even in canonical expositions of the emptiness of things, the idea is often explained in different, and not obviously equivalent, ways. A central part of its meaning is, however, that nothing has intrinsic existence (self-existence). In terms of Western philosophy, we might think of this as saying there are no substances. Nothing exists in and of itself. Everything that exits does so in as much as, and only in as much as, it relates to other things. It has, so to say, only relational existence.[1]

To explain the notion of relational existence more clearly, let us start by looking at a couple of examples from Western philosophy. The first comes from the philosophy of space and time. Famously, Newton was an absolutist about space and time. He held that:[2]

[1] Another part of its meaning is that nothing has intrinsic properties—in the jargon of Western philosophy, primary properties. All properties are relational. Indeed, the claim that existence is not intrinsic can be thought of as a special case of this: even existence is relational.

[2] *Principia Mathematica*, Scholium to the Definitions.

> absolute, true, and mathematical time, of itself, and from its own nature, flows equably without relation to anything external, and by another name is called duration.

and that:

> absolute space, in its own nature, without relation to anything else, remains always similar and immovable.

In other words, Newton held that spatial and temporal locations exist in themselves, and would exist even if there were no physical things that occupied space and time.

By contrast, Leibniz was a relationalist about space and time. He held:[3]

> space to be merely relative, as time is; that ... it ... be an order of coexistences, as time is an order of successions.

In other words, spatial and temporal locations have no intrinsic existence. Physical events bear temporal relations (befores and afters) to each other, and there is nothing more to occurring at a certain time than having certain of those relationships to other things. Similarly, physical objects have spatial relations to each other (norths and souths), and there is nothing more to being in a certain place than having certain relations to other things.

Newton and Leibniz each gave various arguments for their views, but we will not be concerned here with what they were, or who was right in the matter. I have stated their positions only to illustrate relational existence. For Newton, spatial and temporal locations exist in and of themselves: they have self-existence. By contrast, for Leibniz, they have no self-existence. A spatial/temporal position is simply a locus in a field of spatial/temporal relations. That is, it has only a relational existence.

That various entities have no absolute, but only a relational, existence, is a not uncommon view in Western philosophy. Of other examples that could be given, let me cite just one.[4] Words and sentences have meanings. What is the status of meanings? Many philosophers and linguists have held that there are intrinsically existing meanings, entities which exist over and above words and sentences. A notable person who held this view was, of course, Frege. For him, the senses of worlds and sentences exist in just this way.

[3] Leibniz' third letter to Clarke.

[4] Another is Marx's view of the commodity. This is not a thing in itself, but just something that occupies a relational role (notably between producer and consumer) in certain capitalist practices.

By contrast, structuralist linguists, beginning with Saussure, hold that meanings do not have self-existence. Words and sentences enter into various kinds of relationships with other worlds and sentences. For example, 'blue' contrasts with 'red' in a way that 'scarlet' does not. To have a certain meaning is simply to be related to other words/sentences in certain ways. That is, meanings have no intrinsic existence, merely a relational one.

Though I have used meanings as an illustration, it should be clear that I could have used any other example where a structuralist account can be contrasted with a non-structuralist one. For example: a platonist account of mathematical entities, versus a structuralist account. Quite generally, structuralism attributes certain kinds of entities only relational existence: the entities concerned have no self-existence.

Examples such as the previous ones illustrate, I hope, the notion of self-existence versus relational existence. To have self-existence is to exist independently of anything else. By contrast, things that have relational existence are simply *loci* in a field of relations, and are individuated by that location.

3 Śūnyatā

So much for the notion of relational existence. As we see, it is, in fact, a familiar notion in Western philosophy. Now to śūnyatā. Śūnyatā, in the sense we are going to understand it here, is simply the doctrine that *every* entity that exists has relational existence. There is no entity that has intrinsic existence.

I cannot think of any Western philosopher who has endorsed exactly this view, but it is orthodox in Mahāyāna Buddhism. A canonical defence of the view was provided by Nāgārjuna, the second century Indian philosopher, particularly in his text *Mulamadhyamakākirikā*. In this text, Nāgārjuna goes through all the things that one might think to have self-existence, and argues that they do not. Many of the arguments employed concern the kind of thing in question, such as matter, time, consciousness. But some of the arguments are quite general. Here is one such argument from Chapter 5 (or at least, my interpretation of it—interpreting Nāgārjuna is always a sensitive issue).

Take an object that one might suppose to have self-existence. Since the argument is quite general, *anything* will do, but for the sake of illustration, suppose we take Aristotle. Aristotle had various properties: having certain parents, being born in Stagyra, being called '*Αριστωτελες*', and so on. Now, to be Aristotle is to be the bearer of those properties. Any entity which bore (related to) those properties would *be* Aristotle. Aristotle, then, does not have self-existence: to be (identical to) Aristotle is to be related to those properties in that way.

Western philosophers might balk at some of the details of this argument. To be Aristotle, it may not be necessary to bear *exactly* those properties which Aristotle had, maybe just some of them; perhaps just the essential ones, or perhaps just enough to constitute a family resemblance, or whatever. Here we touch upon the details of Western theories of identity. The details are not important, however. On all the theories, to be Aristotle is to be related to a certain bunch of properties in a certain way. In other words, Aristotle has relational existence.

Perhaps the only way to avoid this conclusion is to suppose that there is something to being Aristotle which is independent of the possession of any properties. Maybe Aristotle is the substance which *bears* all the properties. Different substances (or bits of substance?) might bear exactly the same properties; still, one might be Aristotle and the other might not be. This thought does not survive long, though. On this account, to be Aristotle is to be a certain substance, *to be that very substance*. Aristotle's existence is certainly not, therefore, independent of the bearing of properties. On this account, to be Aristotle is exactly to be the bearer of the italicized property. We might call the property its individual essence; this is simply essentialism of an extreme form.[5]

4 Nihilism

What to make of the foregoing argument is a nice point, but it at least makes the view that everything is empty (of inherent existence) an intelligible, and perhaps even plausible, one. Let us suppose that the doctrine is correct. Then for any object, a, a has no intrinsic existence; it is merely a locus in a field of relations, $R_0, ..., R_n$. But what of these relations? If the doctrine is right, these have no intrinsic existence either: each relation, R_i, is simply a locus on another field of relations, $R_{i0}, ..., R_{im}$. And what of each of these relations, R_{ij}? Each of them has no intrinsic existence, but is simply a locus in another field of relations, etc.

Clearly, we have an infinite regress,[6] and the regress may well be taken to be vicious, in the following sense. The existence of any thing is constituted by,

[5] There is another move one might make here. This is to claim that the object and its individual essence are the same thing, and that the relationship involved is simply, therefore, identity, which is not a real relationship at all. Whether or not this is the case, the object/essence bears more than just the relationship of identity to itself. It bears the relation of instantiation. This is certainly a real relation.

[6] The regress may go round in a loop at some stage; but a regress that repeats itself is still a regress.

and only by, the existence of other things, whose existence is constituted by, and only by, the existence of other things, and so on. Since there is nothing that grounds this process, there is nothing that ultimately constitutes the existence of anything. Nothing, therefore, exists. Emptiness entails nihilism.

This argument is also of a kind that is familiar in Western philosophy. Let me give another couple of instances of it to try to render its cogency clearer. The first is from Kant's *Critique of Pure Reason*. In the Second Antinomy, Kant gives arguments for the claim that matter must both be and not be infinitely divisible. In his own words, the *reductio* against infinite divisibility goes as follows (A434=B462):

> Let us assume that composite substances are not made up of simple parts. If all composition then be removed in thought, no composite part, and (since we admit no simple parts) also no simple parts, that is to say, nothing at all, will remain, and accordingly, no substance will be given.

In other words, consider any substance, suppose that it is a composition of smaller parts, and each such part is a composition of smaller parts, and each such part is a composition of smaller parts, and so on *ad infinitum*. Then upon complete decomposition, there is nothing left. Thus there could have been no substance there in the first place.

The second example comes from Wittgenstein's *Tractatus*, and is to the effect that the world must contain what he calls 'substances'. These are not *physical* simples as in Kant, however, but *logical* simples; that is, things which ultimately ground the meanings of propositions. In his own words, the argument goes as follows (2.0211-2):

> If the world had no substance, then whether or not a proposition had a sense would depend on whether another proposition was true.
>
> In that case we could not sketch any picture of the world (true or false).

The argument is a dark one. But what I take it to mean is something like this.[7] Suppose a proposition had meaning in virtue of its decomposition into meaningful parts; and that each of these was meaningful in virtue of its decomposition into meaningful parts, and so on *ad infinitum*. Then there would be nothing, in the end, to determine the meaning of anything. No proposition would, then, have determinate sense.

All three of these arguments just considered have the same structure:

[7] *Beyond the Limits of Thought*, 12.6.

- If the X of a thing, a, were constituted by a's relationship to other things, whose X were constituted by their relationship to other things, whose X were constituted by their relationships to other things, and so on indefinitely, then there would be nothing to ultimately determine the X of a. a would therefore have no X.

In the three arguments, X is existence, substance, and meaning, respectively. In each case, the X in question disappears under ultimate analysis. In each case, we analyse and analyse—until nothing is left.

At this point one might object. The regress does not show that there is nothing left; just nothing *determinate*. We can simply accept that the kinds of thing in question (meaning, substance, existence) are, perhaps surprisingly, indeterminate kinds of things. But this does not show that they do not exist. The *Vimalakīrti-nirdesa Sūtra*, for example, explicitly advises one to learn to "tolerate the groundlessness of things".

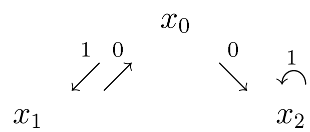
It is here that the debate becomes murky. Can *existence* be indeterminate? What would it be to exist indeterminately? Surely, it might be thought, if something exists, it is exactly what it is? How could there be any indeterminacy about this? But if something is in a continuous transition from being red to being blue, it might be thought, there are times when it is indeterminate as to whether or not it was red. Similarly, if something can come into and go out of existence by degrees, couldn't there be times when it is indeterminate as to whether or not it exists? Wouldn't this be indeterminate existence? The answers to all these questions are opaque.

5 Relational Existence Again

In an attempt to see whether the doctrine of emptiness entails nihilism, let us take another approach. Often, the precise import of an idea can be made clearer if the view in question can be mathematised. This is what I will now do for emptiness: I will produce a mathematical analysis, or maybe better, model, of it. Let us start by returning to the question relational existence. To be empty is to exists only as the locus in a field of relations. How can one understand this mathematically?

A natural way of thinking about objects and the relations between them is as a multi-dimensional graph. Thus, suppose that we have a set of objects, X_0 (why the subscript we will see in due course) and a bunch of relations, ρ, between them. Suppose, for simplicity that $X_0 = \{x_0, x_1, x_2\}$ and that

$\rho = \{R_0, R_1\}$. Then the graph might be depicted by the following diagram:



Arrows denote the relations; superscripts to the arrows denote the relation in question.

This conception takes the objects as independently existing. How could we understand them as simply loci in the network of relations? Suppose that one simply has the relations. Each, R_i, has a collection of instances, $\langle R_i \rangle$. One might think of these as sets of ordered pairs—or actually, ordered triples, the third member indexing the relation. Thus, in the above example, the members of $\langle R_0 \rangle$ are $\langle x_1, x_0, 0 \rangle$, and $\langle x_0, x_2, 0 \rangle$; and those of $\langle R_1 \rangle$ are $\langle x_0, x_1, 1 \rangle$ and $\langle x_2, x_2, 1 \rangle$. Of course, this is already to think of the objects in X_0 as existing in themselves. If one is to jettison this view, one must think of each instance as *sui generis*, and not as a set of pairs.

We may now consider a certain relation, R_L, that holds between instances. Intuitively, the relation holds between those instances that originate from the same locus. Thus, if one again thinks of instances as ordered sets, $\langle x, y, i \rangle \, R_L \, \langle z, w, j \rangle$ iff x is z; but if one is to jettison this idea, R_L must be understood as a primitive relation; and, as the heuristic indicates, it is natural to take this to be an equivalence relation.

One can now identify loci in a very simple matter. A locus is determined by a non-empty bunch of instances that bear the R_L relation to each other. One can, if one likes, think of this as a class. If a is any instance, it determines a locus $\{b : bR_La\}$. The loci are simply equivalence classes of instances.

This is not quite right. Strictly speaking, this captures only those loci that are in the domain (left-hand argument) of some R_i. In general, some loci may be in only the co-domain (right-hand argument) of all R_i; and some loci may be isolated points, and so in neither the domain nor co-domain of any R_i. Thus consider the following diagram:

$$x_2 \overset{1}{\leftarrow} x_0 \overset{0}{\rightarrow} x_1$$

$$x_3$$

Neither x_1 nor x_2 is in the domain of a relation, and x_3 is in neither the domain nor the co-domain of any relation.

The first problem (the co-domain problem) may be avoided by taking every point to be in the domain of some R_i. This is achieved by insisting that ρ is closed under converse relations. Thus, if R is a relation, its converse, R_C, is the relation such that xRy iff yR_Cx. And we require that if $R \in \rho$, $R_C \in \rho$. Call this the *converse closure condition*. Given this, anything that is in the domain *or* co-domain of some relation in ρ is in the domain of some relation.

In the present context, the converse closure condition is a very natural one. If R is a relation on a bunch of objects, R_C is another natural and obvious relation on them, obtained simply by employing the passive transformation. Thus, if R is 'x loves y', R_C is simply 'y is loved by x'.

The second problem, that concerning isolated points, can be avoided by insisting simply that there are none. Again, in the context, this is very natural. If we are taking emptiness seriously, there will be no such points. Every object exists in as much as it relates in certain ways to other objects: if it has no such relationships, it does not exist.

With these two restrictions in place, the construction does indeed capture all loci in a graph. Specifically, and again thinking of objects as independently existing, we can map X_0 into the set of loci by the function that maps x to $\widehat{x} = \{a : \text{for some } y,\ a = \langle x, y\rangle\}$. This is obviously a one-to-one correspondence. Alternatively, and in a standard fashion, one may take $\widehat{x}$ to be simply some particular member of this set. Moreover, any relation, R, on the domain of objects induces a corresponding relation, $R^{\uparrow}$, on the loci. Namely, $\widehat{x}R^{\uparrow}\widehat{y}$ iff xRy. The map $\widehat{\ }$ is clearly then an isomorphism.

What we see, then, is that we may dispense with objects and the relationships between them, and operate equivalently in terms of loci and the relationships between these.[8] The ontology of independent objects may be replaced by an ontology of loci.

6 Emptiness and Nihilism

So far so good. Now let us come to the doctrine that everything is empty. This says that everything has only relational existence. We have just seen how to interpret the claim that the members of X_0 have only relational existence: they are to be understood as loci. Loci are a kind of thing, X_1.

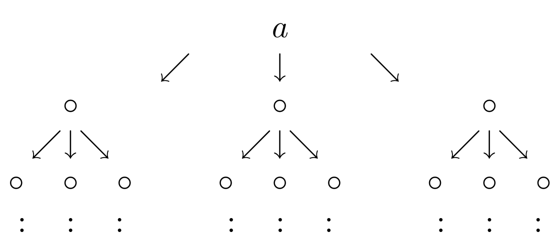
[8] In fact, we may treat not just the objects, but also the relations between them, as themselves loci of a certain kind. Thus consider the relation R_3 between instances, whose intuitive meaning is that $\langle x, y, i\rangle\, R_3\, \langle z, w, j\rangle$ iff i is j. R may now be identified with $\widehat{R} = \{a : aR_3b\}$. If we do this, then the relations on loci may be defined in terms of the relations between the instances, plus set-membership. Thus, let R_P be the permutation relation on instances, such that, intuitively, $\langle x, y, i\rangle\, R_P\, \langle y, x, i\rangle$. Then, as is not difficult to check: $\widehat{x}R^{\uparrow}\widehat{y}$ iff xRy iff for some $a \in \widehat{x}$ and $b \in \widehat{y}$ ($a \in \widehat{R}$ and aR_Pb).

We have seen two ways of understanding X_1. First, it could be a class of sets of relation-instances; alternatively, it could be thought of as a class of (representative) relation-instances themselves.

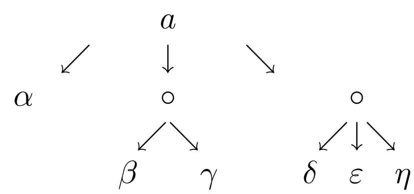
Suppose, first, that we think of loci in the second of these two ways. Then we have traded in an ontology of the objects in X_0 (with relations between them) for an ontology of relation-instances (with relations between them). Now the members of X_1 must themselves be taken to be empty. But since they have a structure of the same kind as that with which we started, we can simply repeat the analysis, to obtain a new set of objects X_2 (with relations between them). But this, again, has the same structure, and is to be analysed in the same way. We repeat the process to the limit (ω times). The result might well be thought to be nihilism. At each stage, a certain ontology is thrown away and replaced by another. In the limit, every ontology has been thrown away.

But the appearance is deceptive. As we have just seen, each X_i can be embedded in (that is, is isomorphic to a part of, X_{i+1}). If we identify the isomorphic objects, we obtain a chain of sets $X_0 \subseteq X_1 \subseteq X_2 \subseteq \dots$. At each stage, what we obtain, therefore, is a richer structure. The result at the limit is $\bigcup_{i<\omega} X_i$. This is the result of taking all objects to be empty, Emptiness itself. And this is not the empty set. Thus, emptiness and is not the same as nothing.

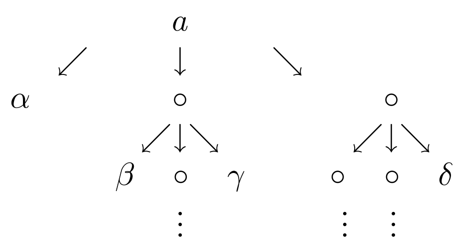
An even more interesting perspective emerges if we interpret loci in the other way, as sets of relation-instances. Again, these instances must be empty, and we can analyse them in the same way, and so in to infinity. To see what happens when we do this, take an object, a, in X_0, as an example. This is a class of instances. Let us suppose, for the sake of illustration, that there are only three instances, b_0, b_1, and b_2. Thus, $a = \{b_0, b_1, b_2\}$. Each of the bs in turn is a class. Suppose, again for the sake of simplicity that each b_i is $\{c_{i_1}, c_{i_2}, c_{i_3}\}$. Then $a = \{\{c_{0_0}, c_{0_1}, c_{0_2}\}, \{c_{1_0}, c_{1_1}, c_{1_2}\}, \{c_{2_0}, c_{2_1}, c_{2_2}\}\}$. And so on. If we pursue this to the limit we obtain a non-well-founded set that we might can be depicted in the following way:



In orthodox set theory (ZF with the Axiom of Foundation) there are no non-well-founded sets of this kind. But there are perfectly respectable set theories where there are such sets.[9] In such a set theory, membership regresses *may* bottom-out to give a perfectly well-founded set. Thus, consider the well-founded set $a = \{\alpha, \{\beta, \gamma\}, \{\delta, \varepsilon, \eta\}\}$, where the Greek letters represent non-sets (or the empty set). This is represented by the diagram:



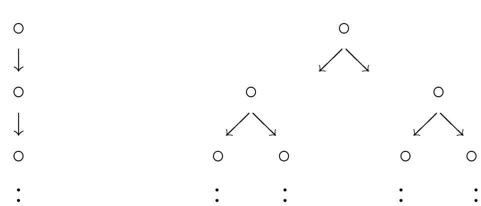
Alternatively, some chains may bottom-out, whilst others do not. Such a situation might be the following:



Alternatively, no branch may ever bottom out. Let us call sets where this is the case *purely non-well-founded sets*.[10] When the analysis of relational existence is pursued to its limit, this is the sort of set to which it gives rise. There is never a ground to the regress. As we noted, the *Vimalakīrti-nirdesa Sūtra* tells us to learn to "tolerate the groundlessness of things". In the case of a purely non-well-founded set, the set has no determinate content: it is pure form. Or maybe better, its content *is* pure form. As the *Heart Sūtra*: form is content; content is form.

[9] For example, Aczel's AFA.

[10] One might wonder how to account for the identity and difference of such non-well-founded sets. A standard answer is that two such sets are identical if they have the same graph. Thus, $\{\{\{...\}\}\}$ is distinct from $\{\{\{...\}\{...\}\}\{\{...\}\{...\}\}\}$ since the graphs of these sets are, respectively:



It might be suggested that taking loci as sets, as we do in this approach, the analysis is still incomplete. We are still left with an ontology of purely non-well-founded sets, with the relations between them, such as set-membership.[11] These would seem to have self-existence (non-relational existence). Should we not iterate the analysis on these? In fact, there is no need. As far as the analysis went, the initial set of objects, X_0, could have been anything. There is no problem about supposing that it contained all purely non-well-founded sets in the first place! In this case, we will have: $X_0 \cong X_1' \subseteq X_1 \subseteq X_0$,[12] where the subsets may be proper; but there is nothing impossible about this, even from the perspective of standard set theory, ZF with the Axiom of Foundation.[13]

What is to be said about nihilism on this perspective? Reality is the totality of all objects. We may, in fact, take this to be X_0. This, being a set of purely non-well-founded sets, is itself a purely non-well-founded set. And this is Emptiness itself. Again, we see that Emptiness is not to be identified as the empty set (which is a well-founded set). As the Mahāyāna Buddhists insist, Emptiness is not a nothing: it has a determinate structure, one of pure form.

Indeed, we may take X_0 to be itself one of the objects in X_0. This gives rise to a new regress, in the form: $X_0 \ni X_0 \ni ...$; but this groundlessness is exactly what one should now expect. Proceeding in this way enforces the thought that Emptiness is itself empty—as Nāgārjuna, that most perceptive of Mahāyāna philosophers, insisted.

[11] Indeed, as we saw in fn. 8, we may take $\in$ (and its converse, $\ni$) to be the only such relation.

[12] '$\cong$' means 'is isomorphic to'.

[13] The natural numbers are isomorphic to the multiples of 4; which are contained in the multiples of 2; which are contained in the natural numbers.