Maximality and Microphysical Supervenience

THEODORE SIDER
Rutgers University

A property, $F$, is **maximal** iff, roughly, large parts of an $F$ are not themselves Fs. Maximal properties are typically extrinsic, for their instantiation by $x$ depends on what larger things $x$ is part of. This makes trouble for a recent argument against microphysical supervenience by Trenton Merricks. The argument assumes that consciousness is an intrinsic property, whereas consciousness is in fact maximal and extrinsic.

**Maximality and Microphysical Supervenience**

Ordinary sortal predicates typically express maximal properties, where a property, $F$, is maximal, roughly, if large parts of an $F$ are not themselves Fs.¹ A large part of a house—all of the house save a window, say—does not itself count as a house. A large part of a cat—all of it save the tail, say—does not itself count as a cat. Otherwise in the vicinity of every house there would be a multitude of houses; in the vicinity of every cat there would be a multitude of cats. The linguistic conventions governing 'cat' and 'house' do not count large undetached parts of cats and houses as cats and houses; therefore the properties these predicates express are maximal properties. Maximality is a kind of *border-sensitivity*: whether something counts as a house or cat depends on what is going on around its borders.

Call all-of-the-house-except-for-the-window "house-minus". House-minus would have counted as a house, had the window never existed.² But given the actual presence of the window, house-minus does not count as a house. So whether something counts as a house depends on more than what that thing is like in itself; whether something is attached to that thing matters too. Therefore, the property *being a house* is extrinsic, or *relational*. Likewise for *being a cat*, and many other properties expressed by sortal predicates of English.

¹ See my 2001a. "Large" is of course vague; moreover, large parts of Fs should be disqualified as being Fs because they are large parts of Fs.

² More cautiously: a duplicate of house-minus not attached to a duplicate of the window (or anything else) would count as a house.
As we will see, these considerations undermine a recent argument given by Trenton Merricks (1998) against microphysical supervenience, the attractive principle that the properties of wholes, in the actual world anyway, are determined by the properties of and relations between their atomic parts (where ‘atom’ means ‘atom of physics’, not ‘partless simple’). Why accept supervenience on the small? Because of the unrivaled success of the physics of the small. Physics and related disciplines have been so successful at explaining macroscopic phenomena that it would take a very powerful argument indeed to undermine our faith in this principle.

This is not to say that failure of microphysical supervenience is impossible. Irreducibly macroscopic properties don’t seem incoherent on a priori conceptual grounds, and there is no apparent reason to accept an a posteriori necessity prohibiting their existence. I therefore construe microphysical supervenience as a contingent supervenience claim, restricted to possible worlds in what David Lewis calls the “inner sphere”: worlds containing no perfectly natural properties or relations beyond those actually instantiated. Compare Lewis’s formulation of the thesis of “Humean Supervenience”, the thesis that all facts supervene on the distribution of point-qualities across spacetime. Lewis allows that Humean Supervenience does not hold in all possible worlds, and so claims only that worlds within the inner sphere never differ unless they differ in their arrangement of point-qualities. To claims of contingent supervenience there is the objection that their truth is not the concern of philosophers. Lewis replies (1986b, p. xi):

Really, what I uphold is not so much the truth of Humean Supervenience as the tenability of it. If physics itself were to teach me that it is false, I wouldn’t grieve....

What I want to fight are philosophical arguments against Humean Supervenience. When philosophers claim that one or another commonplace feature of the world cannot supervene on the arrangement of qualities, I make it my business to resist.

Likewise, my goal is to show that microphysical supervenience is not refuted by any of the commonplace considerations to which Merricks appeals. I do not claim to persuade anyone that irreducibly macroscopic properties don’t exist; that can be done only by a long survey of the history of success of the science of the small, together with remarks about Occam’s Razor or something of the sort. I only wish to show that Merricks’s arguments should not persuade those of us inclined to accept microphysical supervenience that such properties do exist.

The principle of microphysical supervenience would be trivialized if the determining properties of the parts included relational properties like being

---

I do think that it is conceptually incoherent that an intrinsic property of an object, \( x \), depend on intrinsic properties of objects that are mereologically disjoint from \( x \); what is not incoherent is that \( x \) have an intrinsic property that doesn’t depend on the intrinsic properties of the atoms that are part of \( x \).
part of a whole with such and such properties. On the other hand, the principle would be straightforwardly false if the properties of wholes said to supervene included relational properties of wholes: two armies with (intrinsically) similar parts might differ with respect to the property being surrounded. A refined principle emerges, which avoids both triviality and straightforward falsity with a restriction to intrinsic properties:

**Microphysical Supervenience (MS)** Necessarily, if atoms $A_1$ through $A_n$ compose an object $O_1$ that exemplifies certain intrinsic properties, then any atoms like $A_1$ through $A_n$ in all their respective intrinsic properties, related to one another by all the same restricted atom-to-atom relations as $A_1$ through $A_n$, compose an object $O_2$ with the same intrinsic properties as $O_1$.

Assuming 'necessarily' is restricted to worlds in the inner sphere, the principle is, I think, plausible. However, Trenton Merricks (1998, 2001) has argued that it is false.

Consider a person, Mary, who has lost her right index finger. And consider next Martha, a person just like Mary except that she has not lost a finger. Many of us would agree that Martha has an undetached part which consists of all of her save her right index finger; let us call this undetached part Martha-minus. Mary, we may stipulate, is a conscious being. But it is plausible that Martha-minus is not, for Martha is a conscious being, and, Merricks says, we do not want to say that there is more than one conscious being in Martha's immediate vicinity. Moreover, by considerations of symmetry, if Martha-minus is conscious, then so will be thousands of other large undetached parts of Martha; but surely there are not thousands of conscious beings in Martha's immediate vicinity. Being conscious, says Merricks, is an intrinsic property. But Mary and Martha-minus, we may stipulate, are made up of atoms that have the same intrinsic properties and stand in the same restricted atom-to-atom relations—they are "atom-for-atom duplicates". So, since Mary and Martha-minus differ over the intrinsic property of being conscious despite being atom-for-atom duplicates, MS is false.

Some will not accept the existence of Martha-minus, a mere undetached part of the person Martha. But if there is no such person as Martha-minus, then MS is false again. For MS as stated above actually contains two claims: first, that if some atoms compose an object then any atoms with the same intrinsic properties and restricted atom-to-atom relations must also compose an object, and second, that this object must have the same intrinsic properties

---

4 The principle is based on the principle MS from Trenton Merricks 1998. Merricks says 'qualitative intrinsic' instead of just 'intrinsic'; let the latter in this paper be understood as meaning the former. By 'restricted atom-to-atom relations' Merricks means spatio-temporal and causal relations.
as the first. For short: “both composition and intrinsic properties of wholes supervene on atomic intrinsics”. Since the atoms in Mary have the same intrinsic properties and stand in the same atom-to-atom relations to each other as do those atoms that are located in all parts of Martha save her right index finger, if Martha-minus doesn’t exist then the first component of MS, that composition supervenes on atomic intrinsics, is falsified. In what follows I will assume that there is indeed such an object as Martha-minus, since I believe in arbitrary undetached parts (and in fact in arbitrary mereological sums5). Given this assumption, what Merricks’s argument challenges is the second component of MS, that intrinsic properties of wholes supervene on atomic intrinsics.

The argument depends on the assumption that being a conscious being is intrinsic, since MS requires only that intrinsic properties of wholes supervene on atomic intrinsics. (As noted, a non-intrinsic property of a thing, for example the property being surrounded, need not supervene on the intrinsic properties and arrangement of its atoms.) But this assumption is suspect, given what we have learned about maximality. Like other sortal predicates, ‘is a conscious being’ expresses a maximal, extrinsic property. Indeed, the argument seems to assume as much: the premise that Martha-minus is not conscious is plausible precisely because consciousness is maximal.

The following argument seems clearly bogus. “Consider a sufficiently large hunk of matter that contains numerous atom-for-atom duplicates of rocks as parts. Surely those embedded atom-for-atom duplicates are not rocks; the property being a rock is intrinsic; therefore MS has again been shown to be false.” The argument fails because the property of being a rock is not intrinsic. Being a rock is a border-sensitive, extrinsic property. Indeed, the premise that the hunk does not contain a multitude of rocks is justified precisely because being a rock is border-sensitive. Merricks’s argument is no better than this failed argument. The example of Mary and Martha-minus is most naturally taken to show that being conscious is maximal, border-sensitive and extrinsic: whether something is conscious, properly so-called, depends on what external things it is attached to.

Merricks claims that the property being conscious has the “mark” of intrinsicalness: it could be instantiated by a lonely object—an object in a possible world in which nothing else6 exists (1998, pp. 61-62). While it is true that intrinsic properties often bear the mark, the mark is an unreliable indicator in the present context, for some maximal extrinsic properties bear the mark as well. Being a rock could be instantiated by a lonely object; neverthe-

6 Other than its proper parts and things necessitated by its proper parts. The term ‘lonely’ is from Langton and Lewis 1998.
less, a thing can be disqualified from being a rock for the extrinsic reason of being embedded in a larger rock.

I have been sliding between speaking of the non-sortal property being conscious and the sortal property being a conscious being. It would be possible to grant that the first property is intrinsic and still resist Merricks’s argument. Merricks’s argument that Martha-minus is not conscious was that there are not thousands of conscious beings in Martha’s vicinity, but perhaps that is so despite there being thousands of things there that satisfy the predicate ‘is conscious’. The idea would be that while ‘is conscious’ is not maximal, ‘is a conscious being’ inherits the maximality of the sortal predicate ‘is a being’, where the latter means something like ‘is a person’. Martha’s large parts are not beings, and are therefore not conscious beings. This response can even accommodate the intuition that the sentence ‘there is only one conscious thing in Martha’s vicinity’ is true: ordinary domains of quantification typically omit objects that fall under no ordinary sortal predicate. I will continue to argue that ‘is conscious’ expresses a maximal extrinsic property, but will note from time to time how this variant response to Merricks could be developed (the differences between the two responses strike me as underwhelming).

To my mind, viewing consciousness (or being a conscious being) as extrinsic is immediately compelling. But further arguments may be given.

One argument appeals to the plausible analysis of intrinsicality given by Lewis in On the Plurality of Worlds (pp. 59-69). Lewis defines intrinsic properties as those that can never differ between a pair of duplicate objects (perhaps inhabiting different possible worlds). Duplicates are then defined as objects whose parts have the same perfectly natural properties and stand in the same perfectly natural relations. The notion of perfect naturalness is primitive, but the idea is that the perfectly natural properties and relations are the most fundamental intrinsic properties and relations on whose distribution the truth about everything supervenes.

Mary and Martha-minus were in essence stipulated to be made up of subatomic parts with the same perfectly natural physical properties and relations—they are made up of duplicate quarks, electrons, and so on, which stand in the same spatial relations and are related by the same forces. Moreover, physics is micro-reductionist: there are no fundamental properties

---

7 Or defined in terms of an undefined notion of the existence of universals or tropes.
8 All that was directly stipulated was that the atoms of Mary and Martha-minus have the same intrinsic properties and stand in the same restricted atom-to-atom relations. But it is a direct consequence of Lewis’s definition of ‘intrinsic’ that if two atoms have the same intrinsic properties, then their subatomic parts have the same perfectly natural properties and stand in the same perfectly natural relations to each other. Moreover, if “restricted atom-to-atom relations” is to have its intended meaning, it had better follow from the fact that Mary and Martha-minus’s atoms have the same restricted atom-to-atom relations that their subatomic parts have the same perfectly natural relations to each other.
and relations of physics beyond those of subatomic particles. (Certainly Merricks has provided no challenge to this idea.) Therefore, all the parts of Mary have the same perfectly natural physical properties and relations as do the corresponding parts of Martha-minus. So unless there are some non-physical perfectly natural properties or relations that differ between their parts, Mary and Martha-minus are duplicates; and since only Mary is conscious (or, is a conscious being), being conscious differs between a pair of duplicates and hence is extrinsic.

Merricks will likely respond that being conscious is itself an irreducibly macroscopic non-physical perfectly natural property, and that Mary and Martha-minus are therefore not duplicates. (Each is a part of itself, recall.) This claim is coherent since nothing in the concept of naturalness requires that perfectly natural properties be properties of microscopic entities. But the example of Mary and Martha-minus doesn’t force us to posit irreducibly macroscopic mental properties, and there is prima facie reason to resist according this status to the mental. Many of us come to the present inquiry with a belief that the perfectly natural properties and relations are physical, and that all else supervenes (globally) on the distribution of these physical properties and relations. The reasons for this belief are very general metaphysical ones, involving the success of science and Occam’s razor, as mentioned above. Nothing in the example of Mary and Martha-minus undermines this belief; and given the belief, together with the Lewisian analysis of intrinsicality, it follows that consciousness is extrinsic. Of course, I have not really defended an argument for this physicalism. But my goal is not to convince the opponents of microphysical supervenience, only to show how Merricks’s argument against microphysical supervenience can be resisted.

A second argument that consciousness is extrinsic proceeds, not by means of an analysis of intrinsicality, but rather by consideration of what evidence we have for claiming that Martha-minus is not conscious. Merricks’s reason for thinking this, recall, was that we do not want to say that there are thousands of conscious beings in Martha’s immediate vicinity. But what is our reason for not saying that? Our everyday methods for deciding whether something is conscious involve noting how the object in question behaves, and, less practically, checking to see whether it has the appropriate “hardware”—the sort of anatomical makeup that supports consciousness. But the large undetached parts of Martha in question share Martha’s brain, so they have the right hardware. And they will behave in the same way Martha does—they will “speak”, “laugh”, and “behave” exactly as Martha does. And yet we do—properly, I think—exclude Martha-minus from the ranks of the genuinely conscious. Why? The answer seems to be conceptual; it’s a conceptual truth that something counts as conscious (or, as a conscious being) only if it’s the “largest” conscious thing in the vicinity. The concept of being
conscious is a maximal concept. If our evidence that there are not thousands of conscious beings in Martha’s immediate vicinity is not conceptual in this way, then it’s difficult to see what that evidence could be.

But if consciousness is a maximal concept, then that concept has a clearly relational element: the conditions under which it applies to a given object, \( x \), mention objects other than (and mereologically disjoint from) \( x \). In particular, the concept of consciousness applies to a thing, \( x \), only if there are no objects disjoint from \( x \) standing in certain spatial relations to \( x \). Any property whose concept has such a relational element (such as the property of being surrounded) is extrinsic; therefore, consciousness is extrinsic.

A final argument appeals to a principle governing intrinsic properties and a thought experiment involving Martha-minus. Let’s use the term ‘twins’ for things with the same intrinsic properties, without committing ourselves to any particular definition of ‘intrinsic’. The principle of isolation (which is in fact accepted by Merricks⁹) seems true:

if object \( x \) exists in some possible world, then there exists another possible world containing nothing but a twin of \( x \) (along with twins of \( x \)’s parts)

The idea is that the intrinsic features of distinct regions of a possible world are modally independent: one can select a region and “cut away” the rest of the world, leaving the selected region intrinsically intact. Applied to the case of Martha-minus, the principle of isolation yields a possible world containing nothing but a twin of Martha-minus. Note that this twin is not embedded in an object with a right index finger, since we’ve “cut away” the entire rest of the world other than the selected region encompassing Martha-minus. Now for the thought experiment: would this twin be conscious? It seems plausible that it would be. After all, it has a working brain, and given that it is not embedded in a larger object that is clearly conscious, there is no argument available like that in the case of Martha-minus to disqualify it. Merricks might claim that since consciousness is an irreducibly macroscopic intrinsic property, any twin of Martha-minus would not be conscious. But recall the feature of the dialectic I have been emphasizing: I am not trying to change the mind of anyone convinced that consciousness is an irreducibly macroscopic property; I am only trying to show that nothing Merricks has said should convince his opponents that consciousness is an irreducibly macroscopic intrinsic property. It is perfectly reasonable for his opponents to claim that

⁹ Merricks actually uses a slightly different but more general version of the principle, and he uses it in an argument (1998, pp. 69-70) that assumes that consciousness is intrinsic; he therefore concludes that the twin of Martha-minus would not be conscious. The present point is that this is just implausible (what would stop the duplicate from being conscious?); the proper conclusion to draw, therefore, is that consciousness is extrinsic.
the isolated twin of Martha-minus *would* be conscious. But since this conscious object is a twin of Martha-minus, it must share all intrinsic properties with Martha-minus. And hence if consciousness were intrinsic, Martha-minus would be conscious as well. Since Martha-minus is not conscious (or at least, is not a conscious being), consciousness (or, being a conscious being) isn’t intrinsic.

So: I claim that consciousness is extrinsic, whereas Merricks disagrees. It might seem that this is merely a dispute over classification. But that reaction would be wrong (and not only by its presupposition that a “mere” dispute over classification would be unimportant). Merricks goes on to argue that since consciousness is intrinsic and Mary and Martha-minus differ over consciousness, global microphysical supervenience fails. Applied to Mary, the principle of isolation guarantees a world containing nothing but a conscious twin of Mary; applied to Martha it guarantees a world containing a lonely unconscious twin of Martha-minus. These worlds are globally alike at the microphysical level, but differ over consciousness. Moreover, the worlds are in the inner sphere, since we have not *added* any perfectly natural property in moving to those worlds from the actual world. So if Merricks were right that consciousness is intrinsic, then consciousness would not even supervene *globally* on the microphysical, in the inner sphere. And no matter what we think about the classification of properties as intrinsic or extrinsic, this result should be alarming. The restriction to intrinsic properties in the principle of microphysical supervenience was needed because of properties like *being surrounded*, which do not supervene on the microphysical intrinsic properties of the bearer. But whether a thing is surrounded *does* supervene on the microphysical properties of the rest of the objects in the world. In general, anyone impressed by the success of the science of the small will want to say that all (qualitative) properties, whether intrinsic or relational, supervene globally on microphysics (at least within the inner sphere of worlds); any such person must, therefore, resist Merricks’s claims. Since the concept of intrinsicality is bound up with other concepts, in this case modal concepts via the principle of isolation, disputes over intrinsicality are not merely classificational.

In “Against the Doctrine of Microphysical Supervenience” Merricks argued against the response that consciousness is extrinsic; that response, he said, would presuppose a bizarre dependence of consciousness on “irrelevant” factors.\(^{10}\) I have said that whether an object is conscious depends on what else it is attached to. Whether an object is conscious can even depend on the presence of a single atom, for if \(x\) is an atom that is definitely a part of Martha (say, an atom in the center of Martha’s brain), then an object consisting of all

---

\(^{10}\) The objection is on pp. 67-68. The objection is complicated; I reproduce here a simplified version that I hope captures its core. On this matter see also Noonan 1999 and Olson 1995.
of Martha save 𝑥 does not seem conscious (again because we don’t want overpopulation of conscious beings); but an atom-for-atom duplicate of such an object where the atom in the brain is not present does seem conscious. Merricks then claims that since the presence of a single atom seems, intuitively, to be irrelevant to consciousness, it would be better to just give up on the idea that consciousness depends on microphysical matters at all; it would be better to say that consciousness is an irreducibly macroscopic intrinsic property.

But of course the presence of a single atom is not “irrelevant” to consciousness. What the case of Mary and Martha-minus seems to show is precisely that whether a thing is properly called conscious does depend on what it is attached to, just as whether something is a rock, or a house, or a cat, depends on what is attached to that object. Before thinking about undetached parts of conscious beings and duplicates of rocks inside blocks of matter, unreflective people presumably think that no such dependence exists, but after confronting the cases this assurance disappears. It is not surprising that everyday folk are not aware of this dependence given the odd nature of the examples that establish that dependence, but the reaction of the folk to the puzzle cases shows that their concepts exhibit dependence on exactly the extrinsic factors Merricks calls irrelevant.

Merricks is right, however, that it would be bizarre to claim that a single atom could make a difference as to whether a thing is anything like conscious. Surely a single atom cannot make a difference between the full range of conscious experiences I enjoy and having the consciousness of a doorknob! But this is all consistent with consciousness being maximal and extrinsic. Although Martha-minus isn’t literally conscious, she has what it takes intrinsically to be conscious. In particular, she has a working brain. All that disqualifies her from consciousness is a seeming “technicality”: the failure of the maximality condition.11 We can introduce the property of “consciousness*”, which is shared by Martha-minus and Mary alike, which a thing has in virtue of having all that is required, intrinsically, for consciousness. Consciousness* is consciousness stripped of any maximality requirement. The intuition that being attached to a single atom could not make the difference between the consciousness of a person and that of a doorknob is captured by the truth that being attached to a single atom does not affect whether a thing

11 Externalists about consciousness (e.g., Dretske 1995) accept further extrinsic requirements on consciousness, involving the conscious being’s history and/or relation to the environment. Externalists might deny that an intrinsic property like consciousness* would be “consciousness-like”, claiming that any property involving phenomenal consciousness has extrinsic requirements. This issue need not be joined here: even externalists could grant the existence of a property consciousness** just like consciousness but lacking the maximality requirement. My argument would then proceed as before. (Notice further that externalism cannot be used to support Merricks’s argument against microphysical supervenience since externalism directly implies that consciousness is extrinsic.)
is conscious*. (Likewise, a defender of the variant response to Merricks could introduce the intrinsic property of \textit{being a conscious being*}, had by any conscious thing that has all of what it takes intrinsically to be a conscious being.)

Appreciating the existence of consciousness* makes it easier to give up the idea that consciousness is an intrinsic property. Consciousness* is as "real" and "genuine" a property as is consciousness. One could imagine a population of beings otherwise like us who utilize the concept of consciousness* rather than consciousness; such beings would not carve nature at its joints any less than we would, nor would they be missing out on any important feature of the world. Their classification scheme would not be "strange" in Eli Hirsch's (1993) sense. There is a certain role that consciousness plays in our thinking, and though consciousness* doesn't quite play that role perfectly (because of the maximality requirement), it comes pretty close. So if we find it hard to accept that consciousness is not an intrinsic property, we should remember that there is something that is almost consciousness, but \textit{is} intrinsic.

Why do we use the word 'conscious' to express \textit{being conscious} rather than \textit{being conscious*}? In general, why do we use 'rock', 'cat', 'house' and other predicates to express extrinsic, maximal properties? Why do we exclude objects from the ranks of the genuine rocks and conscious beings on the basis of technicalities, merely relational shortcomings? The answer lies in our practices of counting and reference. It is convenient to have manageable counts of rocks and conscious beings. Moreover, since reference, whether by names, demonstratives or descriptions, occurs frequently, if not always,\textsuperscript{12} with the aid of applications of sortal terms like 'conscious being' and 'rock', unique reference requires that these terms express maximal, extrinsic, properties.

Merricks says in \textit{Objects and Persons} (p. 103) that overpopulation of conscious* creatures is just as objectionable as overpopulation of conscious creatures.\textsuperscript{13} But there just isn't anything objectionable about the claim that there are many objects in Martha's vicinity that have a property that is \textit{like} consciousness. There is no \textit{metaphysical} puzzle about how the region could hold a multitude of creatures that are conscious-like, for these creatures share nearly all the same parts in common, share a brain in common, and "think" all the same thoughts. Their near-total overlap ensures that they do not "crowd each other out" (mentally or physically), that their conscious* "experiences" are not objectionably distinct or independent, and so on. The only awkwardness they present is that we don't \textit{say} that there are numerous con-

\textsuperscript{12} Even if one is not a descriptivist about the contents of names and demonstratives, surely descriptions play roles in fixing the referents of such terms.

\textsuperscript{13} Following the terminology of an earlier draft of this paper he calls consciousness* "pseudo-consciousness". I use the present terminology to maintain consistency with Sider 2001a.
scious entities in the vicinity. And this awkwardness is eliminated once we appreciate the maximality restriction on our predicate ‘conscious’: the multitudes are merely conscious*, and hence are, in everyday life anyway, ignored.

Perhaps the awkwardness is not completely eliminated. Perhaps coming to believe in a “mighty host”, as Merricks puts it (2001, p. 95), of conscious* creatures in the vicinity of any conscious creature represents a departure from pre-reflective opinion. I say this is a correct departure, for it enables us to keep our faith in the science of the small. Merricks would have us go the other way. We must sacrifice our belief in the eventual completion of microphysics, all because of the oddness of believing in “mighty hosts” of conscious* beings. So the question is one of trust: do you trust science, or do you trust your intuitions, intuitions that may well be merely the result of semantic constraints of maximality?14

References


14 I’d like to thank José Benardete, David Braun, Eric Funkhouser, Europa Malynicz, Karen Neander, Daniel Nolan, and especially Trenton Merricks for their comments.