Teorema
Vol. XVIII/1, 1999, pp. 139-53

Consciousness, Emergence and Naturalism*

Marcelo H. Sabatés

RESUMEN

En este artículo examino algunos aspectos de la posición emergentista desarrollada por Searle con respecto a la conciencia. En primer lugar discuto las razones de Searle para considerar la relación de emergencia como una relación causal, y argumento que su propia posición se beneficiaría con una noción de dependencia no-causal. En segundo lugar analizo la plausibilidad de la estrategia de Searle para mantener la irreductibilidad ontológica de la conciencia en un marco naturalista, criticando en particular su posición de acuerdo a la cual la subjetividad de la conciencia es solamente el resultado de "la pragmática de nuestras practicas definicionales".

ABSTRACT

In this paper I examine some aspects of Searle's emergentist position regarding consciousness. First, I discuss Searle's reasons for considering the emergence relation a causal relation and argue that his own view might benefit from a notion of non-causal dependence. Second, I analyze the plausibility of Searle's strategy for keeping the irreducibility of consciousness within a naturalistic framework. In this respect I criticize in particular his view according to which the subjectivity of consciousness is just the result of "the pragmatics of our definitional practices".

I take Professor Searle's emergentist view on consciousness as the most desirable stance on this issue. It seems to me that emergentism is, *prima facie*, much more plausible than its rivals. Moreover, I believe his account sheds new light towards the understanding of the problem of consciousness. Still, some uncertainties result from the position and urge us to tackle them. In this paper I shall briefly present Searle's approach to the problem of consciousness (section I), raise some difficulties for such an approach (sections II and IV) and explore some answers to these difficulties (sections III and V). Although some of the answers to be explored might go beyond Searle's emergentism, the central purpose of the paper is to point out some areas of his account that seem to need more articulation.

I. Consciousness, Subjectivity and Naturalism

For many, consciousness is what makes the study of the mind recalcitrant to the naturalistic approaches that dominate contemporary philosophy of mind. The problem of consciousness can be seen as a particularly complex instance of a kind of philosophical problem masterfully illustrated by Roderick Chisholm as follows:

> One kind of philosophical puzzlement arises when we have an apparent conflict of intuitions. If we are philosophers, we then try to show that the apparent conflict of intuitions is only an apparent conflict and not a real one. If we fail, we may have to say that what we took to be an apparent conflict of intuitions was in fact a conflict of apparent intuitions, and then we must decide which of the conflicting apparent intuitions is only an apparent intuition. But if we succeed, then both of the intuitions will be preserved. Since there was an apparent conflict, we will have to conclude that the formulation of at least one of the intuitions was defective [Chisholm (1989), p. 65].

Here are the intuitions generating the puzzle of consciousness: On the one hand, there is probably nothing as deeply embedded in our image of ourselves as the intuition that we are conscious beings and that our conscious life is essentially and ineliminably subjective. On the other hand, we have a more recent, but perhaps equally strong, naturalistic intuition that comes with "our overall 'scientific' conception of the world" [Searle (1992), p. 85]. But how can we incorporate an irreducibly subjective phenomenon in a comprehensive, objective account of the world? There are strategies that deal with the problem by treating (at least) one of the intuitions as an apparent intuition. These constitute, so to speak, strategies of resignation. Since the conflict involves two intuitions, we have two different families of strategies of resignation. The first just "quines" consciousness (Dennett) or reduces it to representational components (Dretske, Tye). The second rejects naturalism (Thomas Nagel) or drastically limits its power to deal with "mysteries" such as consciousness (McGinn). Searle's more heroic emergentist account is a compatibilist account[1]. It is the conflict, rather than the intuitions, that is apparent. And, as it is usually the case with compatibilist strategies, it is much more attractive than its rivals. Searle not only provides a framework in which the conflict becomes apparent but also offers a reformulation of the intuitions that aims to correct their defective traditional formulations. Thus, we can rediscover the mind without moving away from naturalism. Let's see how Searle goes about it.

The naturalistic view includes, according to Searle, all accepted scientific theories, but there are two theories that are particularly pervasive when it comes to explain the world, including consciousness. He writes:

> At least two features of [the scientific] world view are so fundamental and so well established as to be no longer optional for well-educated citizens of the present era; indeed they are in large part constitutive of the modern world view. These are the atomic theory of matter and the evolutionary theory in biology. Of course, like any other theory, they can be refuted by further investigation; but at present the evidence is so overwhelming that they are not simply up for grabs. To situate consciousness within our understanding of the world we have to situate it with respect to these two theories [Searle (1992), p. 86].

The atomic theory of matter can be seen as a stratified theory of reality reconstructed as follows:

A. Reality is constituted by a hierarchy of layers or levels.

B. Each level consists of objects and properties which are characteristic of that level.

C. There is a basic level constituted by (i) those objects that physics recognizes as basic (Searle uses "particle" as an "umbrella" term) and (ii) those properties that physics recognizes as basic (spin, position or whatever).

D. Objects or systems belonging to each non-basic level depend on, in the sense of being constituted by the part-whole relation, objects belonging to lower levels and ultimately on objects belonging to the basic level.

E. Properties or features belonging to each non-basic level depend on, in the sense of being emergent upon, properties belonging to lower levels and ultimately on properties belonging to the basic level.

Claim D is of course essential for naturalism since it precludes the insertion of "alien entities" into the material or natural world (recall, for instance, the emergentist's rejection of *élan vital*). But the claim that is most interesting, and the one that requires careful attention, is E. Searle provides some clarification on how to understand the notion of emergence in this picture, and he does this, for the most part, keeping with the emergentist tradition. When we look at properties or features of a whole or system, we need to distinguish between mere "additive" properties like weight and velocity, which can be deduced or calculated from the properties of the parts, and "emergent" features like solidity and transparency that cannot be deduced or calculated from them[2]. Such features, however, can be explained in terms of the causal interaction of the features upon which it emerges. Unlike additive

properties, emergent properties characterize the different layers or levels of the atomic view. However, we have to be careful to distinguish a naturalist notion of emergence (Searle calls it "emergence$_1$") from a more radical notion ("emergence$_2$"). A property of a system or whole is emergent$_2$ just in case it is emergent$_1$ and it "has causal powers that cannot be explained by the causal interactions" of the properties of the parts of that system. Needless to say, emergence$_2$ is not the kind of property that a naturalist should be willing to accept.

Searle's use of evolutionary theory seems uncontroversial and all we need to say for our purposes is that it naturally adds to the atomic view of matter: it helps to explain how, diachronically speaking, some systems have developed emergent biological and mental properties.

Of first importance for the coherence of the compatibilist claim is the reinterpretation of the intuition involving subjectivity. Searle distinguishes between an epistemic sense of "subjectivity" according to which a judgment is subjective just in case its truth or falsity cannot be settled objectively, and an ontological sense which do not

> refer to an epistemic mode. Consider, for example, the statement 'I now have a pain in my lower back'. That statement is completely objective in the sense that it is made true by an actual fact and is not dependent on any stance, attitudes or opinions of observers. However, the phenomenon itself, the actual pain itself, has a subjective mode of existence, and it is in this sense which I am saying that consciousness is subjective [Searle (1992) p. 94].

Searle also reinterprets the typical naturalistic intuition. The reductive theses that naturalists can accept are very diverse in kind. What is particularly relevant here, is the distinction between ontological and causal reduction. Reduction, Searle says, is a "nothing-but" relation. Applied to properties, *P* ontologically reduces to *Q* just in case *P* is nothing but *Q*. Causal reduction, on the other hand, is a claim that relates causal powers of the emergent property to the ones of its emergence basis. Searle discusses the relation between these two kinds of reduction and his view is the following. For typical emergent properties, causal reduction implies ontological reduction. Thus we (ontologically) reduce solidity to molecular movements in lattice structures, heat to mean kinetic energy, and so on. But this is not the case for consciousness:

> When we come to consciousness, we cannot perform the ontological reduction. Consciousness is a causally emergent property of the behavior of neurons, and so consciousness is causally reducible to brain processes. But — and this is what seems so shocking — a perfect science of the brain would still not lead to

> an ontological reduction of consciousness that our present science can reduce heat, solidity, color and sound [Searle (1992), p. 116].

The conceptual space for a compatibilist solution is particularly narrow. On the one hand consciousness has to keep its *irreducibly subjective character* and thus we cannot reduce it to (or eliminate it in favor of), objective, third person, physical phenomena. Consciousness has to have its own, *sui generis* reality: this seems to be achieved by avoiding ontological reduction. On the other hand it has to be a natural phenomenon *dependent in some sense upon biological processes*, and capable of being studied by objective science — a possibility that arises once we distinguish between epistemic and ontological subjectivity. In what follows I shall look closer at these two requirements, and, in particular, discuss whether they can be met at the same time.

II. Dependence and the Naturalist Claim

Now, what is exactly the nature of the emergence relation between consciousness and neurophysiological features, a relationship that allows us to express a clear dependency without ontological reduction? I take this as a most important philosophical question within a naturalistic picture like Searle's. His answer is that mental states are *caused* by neurobiological processes in the same way heat is caused by molecular motion and solidity is caused by molecular movement in lattice structures: Emergent properties are caused by their bases. But is the solidity of an object *caused* by its molecular structure? Is, in general, a macro-property *caused* by its emergence, supervenience or realization base? Here is an opinion — a majority opinion — to the contrary:

> An individual's possession of blue-eye genes at one time causes the individual to possess blue eyes at a later time. This is a fairly normal case of event causation. But there is no temporal distance between the cup's possession of irregularities at its crystal boundaries and its being brittle. [...] Mereological supervenience involves a simultaneous nomological relation [Segal & Sober (1991), p.10][3].

Searle insists that we have merely a causal relation and he has a reason. Searle seems right when he distinguishes the relation between mental and neurobiological properties from the "closer" relation between moral and natural properties. In the case of *evaluative* properties (or just predicates?), the relation is constitutive and, as such, it is not contingent. Searle says: "[...] the analogy with ethics is just a source of confusion. The relation of macro mental features of the brain to its micro neuronal features is totally unlike the

relation of goodness to goodmaking features, and it is confusing to lump them together." But many defenders of mind-body supervenience (or, for that matter, of supervenience for geological or biological properties) see that relationship as metaphysically or at least conceptually contingent. Many physicalists, as should be clear in the passage by Sober and Segal, tend to think that such dependence stays or goes depending on whether the natural laws ruling in our world stay or go. When most defenders of mind-body supervenience say that "we should not think of the relation of neural events to their supervening mental events as causal"[4] they are not claiming that such a relation is constitutive or conceptual, they are claiming that it is a non-causal, contingent dependence relation.

So far we might describe the disagreement as just a verbal one: "you are using 'contingent dependence' where Searle uses 'causation'". But in an important sense it is more than a semantic disagreement. For we are not willing to grant that for every effect, the relation between the effect and the cause is a relation of emergence. We don't want to say that the accident emerged from the highway's being wet, or that the fire in the building emerged from the government's wanting to destroy the incriminating files. The reason is that, as we saw in the presentation of the atomic view of matter, and as Searle himself seems to acknowledge, emergence is tied to the part-whole relation. And that is not the case for ordinary cases of causation. If we want the notion of emergence to be able to play a substantive role in our naturalistic picture, we need to recognize — and if possible clarify — the distinction between causal and non-causal contingent dependence. Furthermore, there are reasons to think that a notion of supervenience offers the right arena to discuss the mind-body problem, since by comparing different supervenience theses we can formulate and decide some central issues involving the relation between the mental and the physical, such as the kind of properties upon which the mental depends (local vs. global dependence), or the strength of the relation (weak vs. strong dependence, type modal operators involved).

There might be another reason why Searle prefers to consider the emergence relation as a causal one. He seems to think that if we do so, we come closer to solve the traditional problem of mental causation. Still, this will be of little help. For the mental being *caused* by its neurophysiological base is clearly compatible with its being epiphenomenal — in particular, of being causally inefficacious *vis-à-vis* behavior.

None of the two reasons Searle seems to use is compelling. What we need is a notion that allows us to ground the particular idea of dependence that preserves our naturalism in agreement with the atomic theory of matter, and causation doesn't seem particularly helpful. The task of distinguishing between causal and non-causal dependence relations is, however, a difficult

one, and one that has been neglected. In the next section I shall try to sketch a way to accomplish that task.

III. An Attempt to Distinguish Causal from Non-causal Contingent Dependence

The official, mostly implicit account of the distinction relies, as illustrated in the quotation by Sober and Segal, on the synchronic/diachronic distinction. Nevertheless, the absence temporal gap between the "source" and the "result" in simultaneous causation raises a difficult problem for the temporal criterion. How can we resist the conclusion that we have to abandon the intuitive diachronic versus synchronic difference between causation and non-causal dependence? The following strategies are possible (*a*) to affirm that the putative cases of simultaneous causation do not constitute genuine cases of causal relations, (*b*) to affirm that they are cases of causal relations but they are not in fact simultaneous, or (*c*) to accept that simultaneous causation happens but say that while the predominant direction of causation is past-future, this is not the case for non-causal relations (so we still have a possible difference based on the diachronic-synchronic distinction). I take (*a*) to be an *ad hoc* move. As for (*b*), it seems counterintuitive to many; I will discuss it below. Option (*c*) seems the *prima facie* most promising strategy; it can be developed along with an appealing temporal account of causal asymmetry which focuses on the temporal directionality of a causal chain rather than of each of its links[5].

Such an account claims that causation in general involves a chain from C to E involving intermediate links e_1, e_2, ... e_n, such that *C* directly causes e_1, e_1 directly causes e_2, ... , and e_n directly causes *E*. It then claims that the temporal requirement is not something that we have to ask for each adjacent pair of links connected by direct causation. The temporal requirement should be globally imposed to a causal chain with the following two constraints: "(1) require that causes precede their non-simultaneous effects, and (2) maximize causal continuity" [Horwich (1987), p. 136]. This offers an attractive picture to accommodate simultaneous causation.

Now, we could say that while causal chains comply with the Humean time-order requirement, this is not the case for (chains of) non-causal contingent dependence relations. This seems true but probably won't help. For there could be chains which involve (non-simultaneous) causation between many of their links but in which some of their links are tied by non-causal dependence relations. How can we distinguish a chain like this from a chain containing non-simultaneous causation except for some of their links, which are tied by simultaneous causation? And in particular, how can we distinguish instances of direct simultaneous causal relations from instances of non-causal dependence relations? Thus, in spite of containing an appealing ac-

count of the temporal asymmetry of causation, strategy (*c*) doesn't solve our problem.

We just have strategy (*b*) remaining, i.e., to deny that the cases of causation at stake are cases of *simultaneous* causation. In particular, this alternative must show that in the typical examples of putative simultaneous causation the cause occurred before the effect. The iron bar's being hot, for instance, did precede its glowing. This alternative is taken by Beauchamp & Rosenberg in their defense of the Humean condition of the temporal precedence of the cause. Their defense is supported by two related arguments. The first says:

> The assertion that two events are simultaneous requires that there be no temporal gap between them, not even one shorter than detectable by available instruments. This negative existential is no more susceptible of conclusive verification than the opposite claim that there are no cases of simultaneous causation [Beauchamp & Rosenberg (1981), p. 238].

This argument may be seen as unfair since what the defender of simultaneous causation is saying is that in some cases of macro-causation intuition tells us that the cause and the effect properties are instantiated simultaneously. So the burden of proof seems to be on the side of the defender of temporal asymmetry. But a second argument also provided by Beauchamp & Rosenberg may relieve the burden. They say that:

> We have broad and impressive theoretical reasons to doubt that simultaneous causation occur. Atomic theory and the theory of relativity both provide for a lag between [...] the heating and the glowing of an iron bar. [...] the heating of an iron bar involves absorption of energy by its constituent atoms, and its glowing consists in the radiation emitted from their outer-shell electrons. Within such chains of events and their aggregation into detectable glowing there is a vast scope for temporal asymmetry between cause and effect [Beauchamp & Rosemberg (1981), p. 238].

Perhaps this argument is enough to ensure that there is a *temporal asymmetry* somewhere in the process; maybe all other putative cases of simultaneous causation can be accounted for in similar ways. But I am not quite sure that the argument succeeds in guaranteeing a *temporal gap between the two macro events*. In any case, I want to avoid the controversy over whether the macro events themselves are simultaneous or not. This issue seems to depend on how finely we are prepared to discriminate between the times in which macro events occur without doing violence to their nature *qua* macro events. What is significant for me is that in the second argument (and I think this can be applied to other examples) the temporal asymmetry is found as soon as the causal relation is in some way mereologically "extended". And I think that this

is in some way mereologically "extended". And I think that this has important consequences for our purposes, independently of the success of this argument against simultaneous causation.

Some clarification on the concept of extendibility is in order[6]. The notion can be applied to the following two cases. First, given a causal relation, we may be able to provide intermediate links connecting cause and effect properties. Second, given a relation of mereological dependence, we may be able to provide intermediate links between base and supervenient properties. In both cases, we extend by providing intermediate links. However, we may widen the meaning of "extendibility" so as to encompass other cases in which given a relation, we augment our discriminative power by providing a more complete picture of the mechanisms involved is such relation. In this wider sense, when we mereologically decompose a macro relation by providing the supervenience bases of its relata, we are extending the original relation.

Let's see what can be said about the cases of simultaneous causation. In cases such as the heating and the glowing of the iron, the temporal interval in which macro properties are intuitively thought to be instantiated (and macro events are intuitively thought to occur) is probably not fine enough to account for a temporal difference. Macroproperties which are constitutive of macroevents are instantiated in temporal intervals whose beginnings and endings are roughly discriminated by unaided perception. On the other hand, a fine enough temporal discrimination can be made for the case of the micro components into which the causal relation between an iron bar's heating and its glowing can be mereologically decomposed (the absorption of energy by the bar's atoms, the radiation emitted by the electrons). So once we decompose or micro-extend the macrocausal relation we find the temporal gap required for causation. What we have at the macro level is causation, but we have no simple way of distinguishing it from non-causal contingent dependency. So extendibility, in the sense of mereological decomposition, helps us to do the job. It is reasonable to think that in many cases mereological decomposition of a causal relation will give us not just a micro causal relation but a micro causal chain containing (more or less) basic properties as links. In this case each link will be temporally prior to the next.

As we saw, non-causal contingent dependency (mereological dependency in particular) is also extendible. But it is not extendible along the temporal dimension. More precisely, we don't find a temporal gap once we extend the mereological relation. When we extend or interpolate mental-to-physical dependence, for instance, what we cite would be biological and chemical properties which occur simultaneously to the mental and the physical ones. The intermediate links are properties of increasingly simpler parts which are exclusively related by the part-whole relation. On the contrary, the

mereologically simpler properties involved in the extension of a causal relation are temporally ordered. Along these lines we can sketch a proposal for distinguishing causal from non-causal contingent dependence relations:

> (C) A contingent dependence relation is causal iff (i) its source is temporally prior to its consequence[7], or (ii) it can be extended (in the sense of mereologically decomposed) into a relation (or chain) in which the source is temporally prior to its consequence.
>
> (NC) A contingent dependence relation is non-causal iff (i) its source is not temporally prior to its consequence and (ii) it cannot be extended into a relation in which the source is temporally prior to its consequence.

While my proposal may appear close to strategy (b) (the denial of simultaneity), it is compatible with —and perhaps closer in spirit to— strategy (c) (the defense of asymmetry through temporal directionality of causal chains), since the account leaves room for simultaneous causation at the macro level and the notion of extendibility makes the causal chain (at the micro level) the ultimate judge of the nature of the relation.

IV. Causal Powers and Irreducibility for an Emergent Consciousness

Once we develop a satisfactory notion of dependence that can comply with the requirements of naturalism[8] we have half of what we need to defend an emergentist position, and thus half of what we need for our compatibilist solution to the problem of consciousness. The other half is to make sure that we keep consciousness as a real, irreducible feature. According to a widely accepted view, a feature or property can be real only if it is causally efficacious.

Now, there are well-known arguments trying to show that emergent, non-reducible properties are causally inefficacious. Perhaps the strongest among those problems is the so-called "problem of causal exclusion". I will not rehearse here this problem, first formulated by Norman Malcolm, and forcefully developed by Jaegwon Kim. Let me just enunciate the main point: for the mental to be causally efficacious it has to be able to cause behavior or other mental states. But its ability to cause behavior is preempted by the physical (neural) state that surely causes behavior. And its ability to cause another mental state is preempted by the physical (neural) state that is the emergence basis of such a putative effect. Therefore the mental is epiphenomenal: no conscious state is causally efficacious. And this, together with a plausible causal criterion for

property reality yields the conclusion that consciousness is not a real, irreducible feature.

A first possible reply is that since mental properties are just neural properties the problem evaporates. That sometimes seems to be the import of Searle's claim that when we forget about our dualist prejudices we solve the mind-body problem. But this reply doesn't seem to work in the context of our discussion on consciousness. For an important part of our picture is that consciousness cannot be *ontologically* reduced to neural features. It is interesting to note that for all other cases of emergent properties that answer might be enough to block the exclusion problem. Searle accepts that when ontological subjectivity is not involved we *do* have ontological reduction. And if the higher-level features of a system are nothing but other, more basic features of that system, then there cannot be causal competition to begin with. But the case is different for conscious mental properties; they are something over and above their neural sources. It is not so important if we don't want to accept the label of "property dualism". Even if mental properties are just properties of the brain, they are not just reducible to the neural bases. Thus, it seems that we have to accept that, at least *prima facie*, we might get the kind of causal competition that Malcolm and Kim pointed out.

Searle could (and he surely would) reply that we can find an answer to this in his claim that the irreducibility of consciousness "has no deep metaphysical consequences" [Searle (1992), p. 122]. But what does it mean that *ontological* irreducibility has no deep metaphysical consequences? Searle says: "[The irreducibility] merely shows that in the way *we have decided* to carry out reductions, consciousness, by definition, is excluded from a certain pattern of reduction." But recall that for Searle conscious states have to be intrinsic, ineliminably subjective states. It seems to me that if we take this route we are dangerously close to a subtle strategy of resignation: there is no feature in the world that amounts to the subjectivity of consciousness; such a subjectivity is just a matter of "the pragmatics of *our definitional practices*" [Searle (1992), p. 122, *my italics*]. This begins to look like a recognition of the claim that the intuition about the reality of subjective experience is only an apparent intuition and in this direction the mind cannot be rediscovered.

An analogy: How would we evaluate a theory that attempts to solve the conflict between free will and determinism by saying that the reality of free will is just a matter of our definitional practices with "no deep metaphysical consequences"? I suspect that many (not all, of course) would claim that a theory with such a deflationary view of freedom hardly qualifies as compatibilist.

There is, within Searle's view, a further difficulty when we want to claim that conscious states are ontologically irreducible, a difficulty that does not rest upon the exclusion problem. If we accept the principle that in order for a property to be real it has to have causal powers, we surely need to ac-

cept a corollary defending that if a property has no causal powers over and above the causal powers of another property, the first one is identical to the second one. But recall now what we consider to be a reduction:

> The basic intuition that underlies the concept of reductionism seems to be the idea that certain things might be shown to be *nothing but* certain other things. Reductionism, then, leads to a peculiar form of the identity relation that we might as well call the "nothing-but" relation: in general, A's can be reduced to B's, iff A's are nothing but B's." [...] At the very outset it is important to be clear about what the relata of the [reduction] relation are [Searle (1992), pp. 112–3]

Recall also that, as naturalists we need to reject emergence$_2$ and favor emergence$_1$. Emergence$_1$ without emergence$_2$ implies, as we have seen, causal reduction. This means, according to the definition, that the emergent property's causal powers are "nothing but" the causal powers of its emergence basis. And this in turn implies, by the corollary mentioned above, that the "emergent" property is identical to the neural basis, and therefore ontologically reducible to it. Therefore, we either claim that consciousness is causally autonomous and thus emergent$_2$, in which case we seem to abandon naturalism, or we surrender the ontological irreducibility of the subjective character of consciousness. Again, it seems that Searle's emergentism has serious difficulties to articulate a compatibilist position for the problem of consciousness.

V. Ontological Epiphenomenalism Plus Explanatory Emergentism: Compatibilism at a Different Level

It seems that if we take ontological irreducibility seriously, we cannot provide the mental with causal powers. We can give the mental the causal powers of its neural basis but the price seems to be to abandon the subjectivity intuition. We can make the mental causally independent from the neural, but the price seems to be to abandon the naturalistic intuition. Do we still have room for a compatibilist view? I believe we do and that its consequences are not so bad as we might think at first sight. But I also believe that such a view would be, even if successfully developed, much less attractive than a full-fledged emergentism. Here is a sketch of how that view would look like.

We first accept that conscious states are causally inefficacious. Then we need to drop the requirement that for a property to be real it has to have (active) causal powers and replace it by a more liberal criterion. We then relax the connection between emergence and causal powers. We can continue

to speak about emergence since we keep the dependence relation between the mental and the neural, but the view so construed is in fact ontologically epiphenomenal. But mental properties are still intrinsic properties of organisms. We accept that causation just happens at a physical/neural level. Does this mean that consciousness and, a fortiori, mental states, lose any kind of *relevance* with respect to our physical world? Does this mean that we lose any place for explanations involving mental states? Not necessarily. For the step that goes from causal inefficacy to irrelevance (and in particular explanatory irrelevance) is only sound if we grant the view that every explanation has to meet the following causal condition for explanations:

> (CCE) "*E* because *C*" is an explanation only if, provided that *C* denotes *c* and *E* denotes *e*, *c* causes *e*, where *C* and *E* are linguistic expressions and *c* and *e* are events or properties.

If we accept the results of the exclusion argument (or, for that matter, any other epiphenomenal argument) no explanans-candidate involving a mental event or property can be explanatorily relevant since an explanation containing it would not meet condition (CCE). This means that we cannot appeal to mental properties or states to explain either behavior or other mental properties or states. Interestingly, this also implies something that many would find even more puzzling: that our mental states or properties cannot be explained at all. For as I argued in sections II and III mental property non-causally depend upon neural ones. Since mental properties are uncaused, they can never be in the position of "*e*" in the clause "*c* causes *e*" of (CCE). And this means that they are explanatorily isolated. These are consequences of accepting causalism, the view defending (CCE). Still, if we think that there are important dependence relations that are not causal, we might be willing to replace (CCE) by a more pluralistic cousin:

> (DCE) "*E* because *C*" is an explanation only if, provided that *C* denotes *c* and *E* denotes *e*, *e* depends[9] on *c*, where *C* and *E* are linguistic expressions and *c* and *e* are events or properties.

Now, it is obvious that we have the weapons to block the inference from causal inefficacy to explanatory irrelevance. On the one hand, according to (DCE) we can have explanations even if there are no causal relations grounding them. On the other hand, conscious properties are tightly connected to physical properties through dependence relations, so (DCE) makes room for the psychological to be involved in explanations. If we take this picture seriously, the crucial task is to show *exactly* how epiphenomenal properties, and in particular consciousness, can have "emergent" explanatory relevance[10].

In this view, both of our conflicting intuitions are kept. Naturalism is unaltered. Irreducibility is reinterpreted as not involving causal autonomy —not involving causal efficacy, in fact. Is this a price we are willing to pay? Compatibilism seldom comes for free.

Department of Philosophy
Kansas State University
204 Kedzie Hall, Manhattan, KS 66506 USA
E-mail: sabates@ksu.edu

NOTES

* Thanks to the audience of the *IX Seminario de Filosofía y Ciencia Cognitiva*, particularly John Searle and Manuel Liz, for valuable discussion.

[1] It is not surprising that emergentism as a relatively structured philosophical doctrine was developed by Morgan and Alexander as a compatibilist solution to the "life-matter" problem as opposed to two strategies of resignation: vitalism and reductivist materialism (cf. Mc Laughlin (1993)).

[2] Cf. Searle (1993) pp. 111–2. (The term "additive" is not Searle's.)

[3] Similar views can be found, for instance, in Fodor (1987), Kim (1974), (forthcoming), Yablo (1992) and Jackson & Pettit (1990).

[4] Kim (1979) quoted in Searle (1992), p. 125. Two caveats: First, this might be slightly misleading since if we understand mind-body dependence as strong supervenience, it is only the inner modal operator that would be considered as bearing nomological necessity. The modal force of the outer operator would, I think, generate some disagreements among physicalists. Second, it needs to be acknowledged that some influential physicalists such as Kim and Jackson are beginning to favor a stronger-than-nomological relation for the mind-body case in their most recent works.

[5] Horwich (1987) chapter 8, defends a position like this. My remarks below are directed against this theory as a way of differentiating causation from other dependence relations (something which is not among Horwich's purposes), but not against Horwich's account of causal asymmetry.

[6] The concept is introduced by Peter Bieri in his (1992), but his use is restricted to the first case.

[7] We need this clause since extendibility cannot be applied to direct causal relations between basic properties. But this is not a problem for the present use of this notion insofar as we accept that at a basic level there is always a temporal gap between cause and effect.

[8] The proposal presented in section III in just a preface to the discussion of which kind of notion of supervenience or non-causal dependence is appropriate for the mind-brain relation. For some ideas on this, see Pérez & Sabatés (1995).

[9] We would probably want to add "asymmetrically" before "depends".

[10] The explicatoriness of epiphenomena is accepted by Sober (1986) and Bieri (1992). A partial model for epiphenomenal explanations is developed by Jackson & Pettit (1990) and Sabatés (1997)

REFERENCES

BEAUCHAMP, T. & ROSENBERG, A., (1981), *Hume and the Problem of Causation*, Oxford, Oxford University Press.

BIERI, P. (1992), "Trying Out Epiphenomenalism", *Erkenntnis*, vol. 36, pp. 283-309.

CHISHOLM, R. (1989), *On Metaphysics*, Minneapolis, U. of Minnessota University Press.

FODOR, J. (1987), *Psychosemantics*, Cambridge, Mass., The MIT Press.

HORWICH, P. (1987), *Asymmetries in Time*, Cambridge, Mass., The MIT Press.

JACKSON, F. & PETTIT, P. (1990), "Program Explanation: A General Perspective", *Analysis*, vol. 50, pp. 107-21.

KIM, J. (1974), "Noncausal Connections", *Noûs*, vol. 8, pp. 41-52.

—— (1989), "Mechanism, Purpose and Explanatory Exclusion", *Philosophical Perspectives*, vol. 3, *Philosophy of Mind and Action Theory*, Atascadero, Ridgeview, pp. 77-108.

—— (1993a), *Supervenience and Mind*, Cambridge: Cambridge University Press.

—— (1993b), "The Non-reductivist's Troubles with Mental Causation" in (1993a), pp. 336-57.

—— (forthcoming), *Mind in a Natural World*, Cambridge: MIT Press.

MALCOLM, N. (1968), "The Conceivibility of Mechanism", *Philosophical Review*, vol. 77, pp. 45-72.

MC LAUGHLIN, B. (1993), "The Rise and Fall of British Emergentism" in Beckerman A. (ed.) (1993), *Emergence or Reduction?*, Berlin, De Gruyter Verlag, pp. 49-93.

SABATÉS, M. (1997), "Exclusion, Programming and backtracking Dependence: The Fate of Psychological Explanations", presented at the Fifth International Colloquium on Cognitive Science, Donostia.

PÉREZ, D., & SABATÉS, M., (1995), "La Noción de Superveniencia en la Visión Estratificada del Mundo", *Análisis Filosófico*, vol. 15, pp. 181-99.

SEARLE, J. (1992), *The Rediscovery of the Mind*, Cambridge: MIT Press.

—— (1995), "Consciousness, the Brain and the Connection Principle: A Reply", *Philosophy and Phenomenological Research*, vol. 55, pp. 217-32.

—— (1997), *The Mistery of Consciousness*, New York, *The New York Review of Books*.

SOBER, E. (1985), "A Plea for Pseudo-Processes", *Pacific Philosophical Quarterly*, vol. 66, pp. 303-9.

SOBER, E. & SEGAL, G. (1991), "The Causal Efficacy of Content", *Philosophical Review*, vol. 100, pp. 1-30.

YABLO, S. (1992), "Mental Causation", *Philosophical Review*, vol. 101, pp. 245-80.