

CAUSAL EXCLUSION AND EVOLVED EMERGENT PROPERTIES

Alexander Bird

Abstract

Emergent properties are intended to be genuine, natural higher level causally efficacious properties irreducible to physical ones. At the same time they are somehow dependent on or ‘emergent from’ complexes of physical properties, so that the doctrine of emergent properties is not supposed to be return to dualism. The doctrine faces two challenges: (i) to explain precisely how it is that such properties emerge—what is *emergence*; (ii) to explain how they sidestep the exclusion problem—how it is that there is room for these properties to be causally efficacious, given the causal completeness of the physical. In this paper I explain how evolved functional properties can meet both challenges.

1 Introduction

1.1 The exclusion problem for higher level properties

A fragile vase is struck with only a moderate amount of force and breaks. It may be possible to describe its breaking in terms of some micro-structural property of the vase, B. Thus the breaking of the glass is caused by the combination of its being B and its being struck. In explaining the breaking we have not yet mentioned the vase’s fragility. It possesses the fragility in virtue of its possessing B. Its fragility supervenes, relative to the laws of nature, on its being B, which is to say that in all possible worlds that share the laws of the actual world, objects with B are fragile. May we also explain the vase’s breaking by reference to the combination of its being fragile and its being struck?

Although it seems perfectly natural to explain the behaviour of things in terms of their being fragile or sturdy, irascible or docile, conducting or non-conducting, and so forth, such explanations, regarded as causal, face a well-known objection. It seems as if we have two combinations that cause the breaking: ⟨the micro-structure B, the striking⟩ and ⟨the fragility, the striking⟩. So it looks as if the breaking is over-determined. Given the striking, B is sufficient to bring about the breaking. But so is the vase’s fragility. Which is the real cause? This is the causal exclusion problem. The completeness of a causal explanation seems to leave no room for any further causal explanation—complete causal explanations exclude one another. Hence it looks as if we are forced to choose between the supposed causes. And if we are forced to choose, then B wins. Clearly, the vase is fragile in virtue of its being B. B is doing the work if anything is,

and so it rather than fragility is the real cause of the breaking. More generally, we think that a final physics will provide all the explanations of events (or explanations of their chances of occurring) if anything can. While physics may leave some gaps (e.g. in the explanation of why some particular fissile nucleus decayed when it did), it leaves no gaps of a kind that can be filled by other kinds of cause. Thus causal explanations framed in a vocabulary other than that of physics seem to be otiose, excluded by the sufficiency of physical explanation.

Are we forced to choose between B and fragility? We would not be if one of the following were true:

- (i) B and fragility are identical. If they are one and the same, then there is no genuine choice;
- (ii) B is causally downwind of fragility or vice-versa. The throwing of the cricket ball and its striking the vase are both causes of its breaking. But that's no over-determination since the throwing of the ball is itself a cause of its striking the vase—the throwing causes the vase to break *via* the striking.

But neither of these is the case. B and fragility are not identical. Take a piece of very old and dry paper. It too is fragile. But it doesn't possess B. The causal basis of its fragility is something else altogether. So B and fragility cannot be identical. Perhaps fragility is identical with some disjunction combining B and the various other possible causes bases of fragility. But that does not help. That looks to be a very gerrymandered, disjunctive property and far from the sort of property that can be said to be a cause. And even if it could be, that does not dissolve the problem, since that disjunctive property is clearly not identical to any of its disjuncts alone. And so the over-determination problem remains.

Nor can we regard one of B and fragility as causally downwind of the other. Some have argued that B is a cause of fragility. But this will not do. For on the one hand it raises its own problems and on the other it does not answer the question posed. It is problematic since we must suppose the cause and effect are simultaneous, rather than, as in the throwing of the ball and its striking the vase, the cause preceding the effect. The same relationship as that between B and fragility holds between Aloysius being 2m and his being tall. The latter is true in virtue of the former. It supervenes upon it, but is not identical to it. But we would not say that being 2m is a cause of being tall. Consequently it is difficult to see being B and being fragile as distinct links in a chain leading to the vase's breaking. And so even if we were to grant that being B causes fragility, we do not regard B as causing the breaking *via* its causing the fragility. From the mirco-structural point of view, B causes the breaking directly. The fact that it also causes the vase to be fragile is an epiphenomenal effect.

1.2 Is overdetermination a problem?

What forces the choice on us is the presumption that overdetermination is impermissible—effects can have more than one set of synchronic complete causes. A common response is we should not be worried about overdetermination when B and fragility have the sort of relationship that they do have. For example, let us imagine that the counterfactual

account of causation is correct. Thus the fragility of the vase is a cause of its breaking, for if it had not been fragile, it would not have broken. But notice that had the vase not been fragile, it would not have had B. Hence we do not have the sort of overdetermination that is worrying. Overdetermination is worrying when X and Y are both held to be causes and were X not to have existed Y would still have existed and caused the effect on its own. If X and Y are fully independent in this way, then there is something worrying about the supposition that they both cause the effect. It is tempting to think that only one of them is the real cause whereas the other is not actually the cause but is merely a backup, something that would have been the cause in the absence of the first.¹ But this sort of worry cannot arise when X and Y are not independent. Clearly there is no worry in the case where X and Y are identical. In that case X cannot be a mere backup to Y's real cause: had Y not occurred neither would X have. The worry also seems to fail to arise when there is a relationship of supervenience. As we saw, had the vase not been fragile, it would not have been B. Being B cannot be a mere backup to the real cause, the fragility. Such is the sort of response articulated by Barry Loewer (2002: 656-657).

The principal reason given by Jaegwon Kim (2002b) for being suspicious of this response is that we should not be too quick to adopt the counterfactual account of causation. As is well known it suffers from numerous objections, and even its repairs do not seem immune from counterexamples.² That fact suggests what one might have suspected in the first place, which is that the counterfactual account is insufficient to capture our intuitive notion of causation. Causation is what makes things happen, or, in Kim's terms, is 'production', or 'generation' (2002b: 674). But a counterfactual relationship does not seem enough to capture those ideas. A counterfactual relationship might be a symptom of causation, but should not be mistaken for it. Thus if B is what causes the breaking, then fragility cannot. The relationship between fragility and the breaking is a shadow of the real causal (productive) relationship between B and the breaking. As a shadow it inherits some of the form of the real thing, viz. the counterfactual relationship with the breaking, but that isn't causation itself.

There is a further reason that one may adduce from Kim for employing a counterfactual account of causation to deflate overdetermination worries. Kim thinks, in effect, that the overdetermination worry is rather more like the two assassins case. Let us imagine that in the actual world a neurophysiological state or event N realizes a mental state or event, M. Both N and M cause a physical event E. According to Kim, had N been absent, M would have still caused the physical effect, E. Loewer's response is that this suggests that compared to the actual world world w_2 , where M occurs and E occurs, but E has no physical cause, is closer than w_1 , where M occurs and is realized by some different neurophysiological state, N*, which causes E. Lower think that this is wrong, and that since we accept "if M had occurred but had not been realized by N it would have been realized by another neural state", we regard w_1 as closer.

I think, however, that Kim's view has more going for it than has been appreciated.

¹Some philosophers think that such cases are not worrying in unusual cases, such as the two assassins case, but do regard ubiquitous overdetermination of this sort as objectionable. (In the two assassins case, two assassins independently fire at the intended victim.)

²Kim's particular objections are that the counterfactual account does not distinguish causation from epiphenomenal relations and pseudo-processes.

Let us distinguish as Lewis does between sparse and abundant properties. The sparse properties are the *natural* properties of the world whereas abundant properties are mere reflections of our predicates. There is an ontological difference. Sparse properties are parts of the world, and the most straightforward reflection of this is to regard them as universals, whereas merely abundant properties are not genuine entities at all. We may ask then, is M (or the property of being in state M) a universal, a natural, sparse property, that has ontological being? Let us first consider the response that it is not. Then ‘S is M’ is a mere predication of S rather than the attribution of a universal to S. The properties we are considering here are either dispositional ones (such as fragility) or, possibly, higher order functional role properties (the property of possessing some physical property that has the causal role ...). Either way, Loewer is right that w_1 is closer than w_2 . The nearest world in which the same functional role predication is true of S, but S is not N, is (trivially) a world in which some other physical state, N^* , realizes the functional role. Similarly, if the dispositional predication is true, we may ask in virtue of what the predication is true. *Ex hypothesi* it is not a dispositional universal. The only suitable candidate would seem to be some new physical state (N^* again) that realizes the dispositional predication. But that success is hollow for the non-reductive physicalist, since this response, that M is a merely abundant property, is equivalent to eliminativism. It permits mental predications alright denies that there are genuine mental properties. So we must consider the other option, according to which M really is a genuine property (a universal). We are now asked to consider the nearest world in which S is still M but is not N. The question now is whether we should expect that world to contain N^* ? It is at least rather less clear, since neither of the reasons given above hold. We asked, in virtue of what is true that S is M? Now we have the straightforward answer ‘S instantiates the universal M’, and so we are not obligated to find the answer in N^* . It may yet be that the nearest world where S is M but not N is one where S is N^* . That would mean that there M has the characteristic that it may be instantiated only when some other kind of property is instantiated. What kind of relationship between universals would require that? The obvious answer would be when one or other of the universals is complex and the other is a part of it. But M is clearly not a conjunctive universal of which N, N^* etc. are parts—if it were, M would always be realized by all its (incompatible) realizers simultaneously. Nor can we regard each N, N^* etc. as possessing M as a part—and even if we did that would not show that we could not have M without one of its realizers. What would work is if M is a disjunctive universal, the disjunction of all its possible realizers. However, there are good reasons Armstrong (1978: 19-22) for denying the possibility of disjunctive universals.³ We need therefore some other reason for thinking that the universal M can be instantiated without having some realizer. Loewer does not provide such a reason. (I shall later argue that for *some* higher level properties, such a reason can be given.)

It looks therefore as if we can talk about causation at a higher level of description (the level of familiar objects and their properties) only when there is the strongest kind

³Kim’s (2002a: 642) view is that M is a disjunctive property. He remarks that “If you would rather reserve “property” for causally/nomologically homogenous properties, that would be all right—we could call M and other such disjunctions something else. But this is too weak a response. For the issue is whether M is a genuinely natural, sparse property, a universal. If it is not, then we have eliminativism, not reductionism, and calling it something else will not help.

of reduction to a lower level, viz. identity. In the absence of identity there is causation at the level of the causal bases, but not at the level of macroscopic properties. Furthermore, the macroscopic properties are not genuine natural, sparse properties. This fits with (but does not entail) a view according to which our macro-level descriptions reflect a view of the world that although not entirely divorced from its natural processes, is nonetheless in many respects shaped by human interests and perspective.

Such a conclusion may not be too disturbing when it comes to questions of whether the fragility of the vase really caused it to break.⁴ Indeed, I accept for many higher level dispositions, such as fragility, the conclusion of Elizabeth Prior, and Prior, Pargetter and Jackson, that they are causally inert. But that conclusion is more disturbing when it comes to other macro-properties and their causal roles. When it comes to mental states and properties, it is very difficult to give up the idea that they are causally effective (not least because philosophers find it so natural to give accounts of mental concepts in causal terms). The only way to avoid the problem would be to adhere to a type-identity theory of mind. But that is now almost universally rejected also, for reasons already discussed.

In what follows I wish to promote a view which argues that under certain circumstances the higher level property can be regarded as natural and causally efficacious while admitting the strength of the forgoing arguments. The idea is that where the higher level properties have been selected for by a natural process, they can be regarded as natural themselves and causally efficacious. In Agustin Vicente's words, that process converts dispositions into teleological functions (Vicente 2002, 2004).

2 Evolved emergent properties

Consider an old, dry, brittle leaf, which crumbles underfoot. The discussion so far leads to the conclusion that the brittleness plays no part in the causal explanation of its crumbling. This is not to say that the leaf is not brittle—it clearly is. Rather it is to say that brittleness is not the sort of natural property that can play a part in causal explanation. Brittleness is thus somewhat like grueness. Things can be grue or not grue. But grueness is not itself a cause of anything, because grueness is not a natural property. If we were ever tempted to say that x 's being grue caused something, it would really be x 's being green, or x 's having the causal basis of green, that is the true cause (or, *mutatis mutandis*, x 's being blue etc.).

A natural way to think of what is going on is to say that there is no universal of grueness or of brittleness. Natural properties are real entities—universals—but unnatural, constructed properties are not real entities. The former, being real, can be elements in causal and nomic explanations; the latter, not being real, cannot explain anything. While the realist about universals has a clear way of explaining why natural properties are causally efficacious and unnatural ones are not, any other acceptable view of properties ought to be able to make a similar distinction.

⁴There are others who reject the causal claim in any case because of Molière's 'dormitive power' objection. I do not.

2.1 Emergent properties

The problem we face is that of finding a way of regarding mental and other higher level properties as genuinely natural and hence causally efficacious. Such properties will be distinct from lower level properties—if they were identical there would not be a problem to face, but as we have seen, the properties we are interested in are not identical to any lower level properties. Genuinely natural, causally efficacious higher level properties that are not identical nor reducible to lower level properties are *emergent* properties. More precisely I take emergent properties to have in common the following features:

- (i) there are genuine, natural properties, with the causal or nomic efficacy that non-emergent properties have;
- (ii) they are not identical with nor composed out of the fundamental properties of physics;
- (iii) they are properties of physical entities;
- (iv) they nomically supervene on the distribution of physical properties.⁵

The role of (iii) is to rule out Cartesian dualism about substances as a form of emergentism. If there were immaterial substances such as minds, their properties would satisfy (i) and (ii). Feature (iv) rules out dualism about properties, the view that in addition to physical properties of things (their charge, mass, etc.) there are further properties (e.g. mental ones) that although possessed by physical entities (e.g. human bodies) are nonetheless fully causally efficacious and not dependent on the physical ones, offering a parallel sources of causal influence. I do not wish to count such properties as emergent, although this is contentious.⁶ (In any case I do not see how there could be properties satisfying (i)-(iii) but not (iv).)

There are two challenges for emergent properties. The first is to explain how they arise: how do they ‘emerge’ from physical properties? We should note that there may be more than one kind of emergent property—more than one way to be emergent. Even so it is difficult to find *any* convincing way of being emergent. The second challenge is to explain how is it possible for emergent properties to exist at all. For as long as physical supervenience holds, viz. (iv), it looks as if whatever the emergent property claims to be doing is really being done by something else at a lower level.

Emergent properties are sometimes thought to arise from the complexity of certain systems in which higher level patterns arise (such as, in chaotic systems, the patterns that arise over the longer term around Lorenz attractors). But that fact alone does not justify a belief in emergent properties. The fact that the patterns are not to be found at the lower level does not mean that they are truly causally efficacious at the higher level. They may be mere epiphenomena, and indeed it seems that they are, for there is no good reason to suppose that they have an causal status of their own of a kind different from the brittleness of the leaf, which is also a property that arises only at the higher level.

⁵This is a characterization of *ontological* emergence. There are also doctrines of epistemological emergence, e.g. Batterman (2001), Clark (1996, 2001). There may well be epistemological emergence without ontological emergence, and maybe also vice versa.

⁶My reason for rejecting parallel non-physical properties as emergent is that they and physical properties would not show the asymmetry that is implicit in the term ‘emergent’.

So mere complexity is not enough. Other have suggested that emergence is to be related to biological phenomena, and this I think is correct. But the important thing is to explain *how* it is that biological phenomena can lead to emergence.

2.2 Evolved properties

Phenotypic properties are clear examples of higher level properties that supervene on lower level properties. And among the phenotypic properties we can include the mental properties of animals, including humans. Being bipedal, having the capacity to see, and being able to engage in abstract thought, are all phenotypic traits of normal adult humans. Such capacities supervene on lower level traits that are the musculo-skeletal, visual, and cerebral structures of man. These are also multiply realizable capacities. Flamingoes are bipedal in a different way from humans, bees see, but employing very different structures, and Martians probably engage in abstract thought employing a different kind of brain. So the arguments above ought to come to the conclusion that these are not genuine natural properties. Such a conclusion would be less easy to accept than the conclusion that fragility is not a natural property.

Consider a bee which flies around a tree trunk and a cat which walks around it. In both cases we want to explain the behaviour by saying that the creature sees the tree. We are inclined to think that this is a more unifying kind of explanation than that which says that the vase and the ancient manuscript were easily destroyed because of their fragility. The unity does not come from an unity in the underlying causal basis—there is none, since the mechanism of vision in bees is very different from the mechanism of vision on mammals. Nor does it come from the fact that both cases can be subsumed under a single disposition or capacity, since the two instances of fragility can be thus subsumed. The unity comes, I suggest, from the similarity of the causal stories behind the existence of the capacity. The causes of the fragility of the various instances fragility of fragile items are various and are typically merely the different causes of their causal bases. However, when it comes to vision, we think otherwise. The immediate causes of vision in that bee and that cat are different—the differing genotypes of the two creatures. But the more distal cause, which is the same as the general cause of the presence of vision in bees and cats, is much the same. It is the selective advantage that vision gives creatures that possess it over those that do not. The crucial thing about selection is that it selects for dispositions, not for their causal bases, or, as Alexander Rosenberg (1994: 25) puts it, ‘selection for function is blind to structure’.⁷ Individual instances of a predator catching its prey may depend on the causal bases of their individual speeds, acuity in hearing and vision, possession of claws or camouflage; many such cases at a certain time may be considered as just so many cases of fragile things breaking, with no especial underlying unity. But when we consider the prevalence of the trait in subsequent generations it is no longer possible to explain matters in terms of the causal bases alone. For the existence of the trait has only its function to thank. The process of selection would have selected the same trait even if underwritten by a different causal basis. Indeed, it does in fact happen

⁷Agustin Vicente (2004: 305) makes a similar point, concerning explanations of the existence of an organ, such as the heart, in terms of its function: “The process of selection did not take its intrinsic properties into account, only their effects”.

that the same trait in a single species is underwritten by more than one causal basis. Codons (sequences of three bases in a strand of DNA) code for the creation of amino acids. Each codon codes for a single amino acid. But since there are more codons than are needed, one and the same amino acid is coded for by several distinct codons. Consequently, the same protein can be coded for by different sequences of bases, and so natural selection is indifferent to such differences in the DNA. At a larger scale differences in protein brought about by more significant differences in the DNA may nonetheless fulfil the same function equally well. Consequently both proteins (and both kinds of genetic material) may be found in a population.⁸

It should be remembered that selection processes do not take place only on genetic material. Neural networks also evolve by selection processes. The animal brain has a high degree of innate structure. But that structure is modified by stimuli under a regime of selection. Those modifications include the development of general capacities (e.g. the capacity for abstract thought) and the results of their exercise (e.g. the belief that root two is irrational). Thus for the same reasons we should not expect the structures underlying the same mental capacities and mental states to be the constant across different individuals. (Such properties (the capacities and states) are thus doubly evolved, in that they have developed by a selection process from an innate structure that is itself an evolved entity.)⁹

2.3 Evolved properties are emergent

The proposal I am making is that such properties should be regarded as emergent. Let us consider such properties against the criteria for emergence I specified above:

(i) Emergent properties must be genuine, natural, nomically or causally efficacious properties. It is *prima facie* highly plausible that evolved capacities and states are indeed natural properties that are causally efficacious. The properties are generated by a natural process. This is what distinguishes them from non-natural properties such as fragility. Of course, it is contended whether they can be causally efficacious—that is the thrust of the exclusion problem. I shall return to this below.

(ii) Emergent properties are not reducible to the properties of physics. The conclusion of the preceding section is just this. We knew all along that mental and (other) biological properties are multiply realizable. The point of the evolutionary story is to remind us why this is so while at the same time making the case for their being natural.

(iii) Emergent properties are properties of physical things. Emergentism is not a form of substance dualism. And the properties in question are properties of physical objects—creatures made of non-vitalistic molecules and nothing more.

(iv) The distribution of emergent properties nomically supervenes on the distribution on physical properties. Strictly, this account may not satisfy a synchronic supervenience claim. For I have characterized evolved emergent properties in aetiological

⁸Such cases are the exception rather than the rule since beneficial mutations are rare and so the chances of two or more mutations at the same locus that are equally beneficial are very small.

⁹Robert van Gulick's (van Gulick 1993) approach to mental causation (cited in this context by Vicente (2004: 307)) develops just this feature of the mind. As Vicente puts it, "the brain ... recruits physical structures according to their causal capacities, and because they have such causal capacities, and then *selectively activates* them."

terms—how they have evolved. An evolved biological system could have a molecule-for-molecule duplicate that has not evolved, but has come about by accident (e.g. swampman). On an aetiological account of emergence this duplicate system would not have emergent properties. Thus supervenience fails. However, the point of the supervenience claim was to rule out entirely parallel non-physical causal properties, and the current proposal does not introduce them. The supervenience claim can be interpreted more locally: there is no possibility of changing *this* system's distribution of emergent properties without changing its physical properties. Or it can be given a diachronic element: no two extended histories can differ in their final distribution of emergent properties without differing in their physical properties at some point.

A different approach to reconciling supervenience with a evolutionary account of emergent properties would be to provide an alternative to the aetiological account of emergent properties. What is important for the current story is that emergent properties are functional. The accounts of biological function are typically aetiological, which is why this story is aetiological also. However, we may account for biological function in terms not of historical evolutionary benefit but in terms of current fitness provision (Walsh 1996). Thus we may argue that swampman does have organs with functions precisely because *he can* reproduce and those organs help him do so. Swampman may thus also be attributed with emergent properties. On this picture emergence is synchronic product of a very particular kind of complexity, the degree of complexity that permits reproduction and evolution. Such complexity has itself hitherto only ever been the product of evolution, but in principle need not be (and recent developments in bio-engineering suggest that the first purely artificial bacterium is not far off).

2.4 The structure of emergent properties

It is worth noting that when we ascribe causal efficacy to emergent properties, we do so in relation to one another as well as to purely physical, non-emergent properties. emergent properties do not emerge singly but in relation to one another. Not only are the functional properties of our bodies and their parts causally related to one another, they are related to the functional properties and capacities of other organisms (most obviously in the case of symbiosis). Equally, our mental properties are responsive to one another—and to the mental properties of other creatures. If prey is faced with a predator and is selecting between fight or flight it is the other's capacities, the ability to fight or to run fast, that is being assessed, as is also the other's mental state (e.g. willingness to fight). Language gives us the ability to respond (mentally) in a highly sensitive manner to the mental states of other. Just have physical capacities co-evolved, so have mental ones.

The picture of the structure of natural properties we should have is this. At the bottom are the fundamental physical properties. Built out of these are the non-fundamental physical and chemical properties, to which strict type-reductionism applies. However, once we get to biological properties, we find a new level of quasi-fundamental properties, the emergent properties. They are fundamental to the extent that they are irreducible, but non-fundamental in that some supervenience thesis holds of them. If there is more than one fundamental physical property, then those fundamental properties form an interrelated family which are responsible in concert for the non-fundamental

(but reducible) physical properties. Similarly, the quasi-fundamental emergent properties form a family that can potentially form the basis of further properties, reducible to them.

Another way of looking at this is in terms of the Ramsey sentences required to describe truly a certain area of science. We might describe some area of chemistry or non-fundamental physics. If so we will describe the relevant properties in functional terms—a property is characterized in terms of its causal relations to other properties. However, we need not regard the functional characterization as describing the property's essence. For investigations at a deeper level will show how the relevant property is reducible to those at that deeper level. However, when we get to the fundamental level, there will be no further reduction. The fundamental properties will have a functional characterization, and that is all that there is to be said about them. According to the dispositional essentialist, in the special case of a fundamental property, its functional characterization *does* describe its essence (for the other properties, their reduction via type-identities describes their essence). Now let us consider the properties quantified over in a true biological or psychological theory. In such cases, there may be some such properties that are reducible to further biological and psychological properties. But just as in the chemical-physical case, we reach a fundamental level whose properties are not type-reducible. As in the physical case, those properties have functional (or dispositional) essences. Only in this case the properties are emergent, not 'utterly' fundamental. But their quasi-fundamentality does mean that they could supervene on an entirely different base, and in that sense, they are an independent set of properties.

3 The causal efficacy of emergent properties

So far it looks as if biological, including mental properties are emergent. While they look to be causally efficacious, the challenge remains to show how they really can be, in the light of the causal exclusion problem.

I shall take as the key to causal efficacy the ability of a property to raise the chances of the supposed effect. In general, events or facts are causes of other events or facts when they raise their chances of occurring: smoking causes cancer because it raises the chance of cancer, hard work is a cause of success because it raises the chance of success. When we analyze the antecedent facts in question, the complex or simple properties we find will be causally efficacious properties.

3.1 Emergent properties as chance-raisers

Biological and mental properties have this chance raising character: the mottled colouring of a moth raises its chance of survival; infection with *Cryptosporidium* raises the chances of diarrhoea; desire for X raises the chances of actions that bring about X, and so forth. That much seems clear. However, our problem is to show on the one hand that the causal work is not really being done elsewhere and on the other hand it is not either overdetermining or, worse, in conflict with chance-fixing determined elsewhere.

Thus one proposal would be that in each case of alleged biological or mental causation there is a physical event doing the causing. In the case of desire raising the chance of action, it is some physical state of the individual that raises the chance. Note first that we do not need to think of these as being in competition, since many factors can raise the chance of the same event: S's smoking raises the chance of S's getting cancer. But then so does S's being exposed to high doses of radiation. Secondly, we note that facts about chance-raising relate, usually implicitly, to a reference class. There is a certain class of people who are disposed such that if they were to take up smoking (perhaps under duress) that they would so compensate in leading healthier lives in other respects, that for them the chances of cancer are reduced. But when we say that smoking causes cancer we are not thinking about that reference class but about the class of normal adult humans. Now let us consider the claim that the chance-raising (of action A) achieved by S's mental property, M, of desiring X, is really achieved by some relevant physical property P of S. We know that not only is M multiply realizable but also that P need not correspond to M in other individuals. Consequently the reference class for which P is a chance-raiser for A may be very small, perhaps limited to S alone. By contrast, M is a chance-raiser for a much wider reference class, that of all adult humans. That shows that S's being M is a distinct chance-raising fact from S's being P. Furthermore, the class consisting of S alone is not a natural reference class. Bogus chance-raising can be achieved by considering unnatural classes (consider the class consisting of (a) non-smokers who are obese, take no exercise, suffer from many infections, and are exposed to high doses of radiation; and (b) light smokers who lead an otherwise very healthy life).

One might instead take the physical state in question to be a disjunctive state, the disjunction of the relevant physical states for all people with M. This seems to give us a large and natural reference class—all humans. This proposal faces an obvious problem, that we should not regard the disjunctive property as a genuine natural property. Gerrymandered disjunctive properties can be made to be chance-raisers even in large natural classes.

So the causal work of emergent properties is not done by anything else. Does that not mean that it duplicates the work of the physical properties or even potentially conflicts with it? That there is no potential conflict is ensured by the fact of supervenience. There will be some *apparent* conflict, since the chances of some physical outcome (the result of an action) may differ when conditional on S's being in M and S's being in P. But that will be because we are implicitly using different reference classes (all humans when considering M, a much smaller class when considering P, perhaps S alone). As the classes converge, the chances will converge. As we may have expected, the real challenge is that the emergent and the physical properties overdetermine their effects.

3.2 Causal exclusion

The concern presented by the challenge is that the complete causal efficacy of the physical properties excludes any genuine causal role for the emergent properties. The attraction of reductionism is that it just eliminates the exclusion problem. Other views that are not manifestly reductionistic may nonetheless trade on its attractions. Thus one might claim to be non-reductionistic by virtue of denying type identity between

the mental and physical yet gain the benefits of reductionism by admitting token identity. But this is misleading. The identity of facts, states-of-affairs, depends on the individuals and properties composing them. If S's having mental property M is the same fact or state-of-affairs as S's having physical property P, then if M and P are genuine natural properties, they must be the same natural property. Conversely, if M and P are distinct properties, then S's having M and S's having P are distinct states of affairs. Such a problem does not arise if the term 'M' does not name natural property. The proposition 'S has M' then may pick out the state-of-affairs that is S's having P, even though M and P are not identical. But in that case there are no genuine, natural mental properties at all. Compare s's being grue. If grueness were a genuine natural property it would be different from the natural property of being green, a's being grue and a's being green would be distinct states-of-affairs. However, as it is grueness is not a genuine natural property, and if we think of 'a's being grue' as picking out a state-of-affairs, or being made true by a state-of-affairs) before the switching time *t*) then that state-of-affairs is just a's being green. That does not amount to a non-reductive view of grueness. If anything, it is a species of eliminativism, since grueness of eliminated from the catalogue of genuine properties. The only difference between 'grue' and 'M' on this picture would be that 'grue' has a straightforward definition, whereas we suppose that 'M' need not. But that is a ground only for denying conceptual reduction, not ontological reduction (or elimination).

Similar remarks may be made about the claim that mental properties are higher-order functional properties—the property of having some physical property with the causal role This is supposed to be physicalistic since the instantiation of the high-order property just is the instantiation of some physical property. But if that is right, then there is no genuine higher-order property in existence. To be sure, we can introduce a higher-order functional *predicate*. But that provides no guarantee that the predicate corresponds to a property any more than defining 'grue' does. As it stands, such a view is physicalistic alright, but it has done nothing to show that there are any higher order properties that fail to be reduced. Its attractions are therefore the real attractions of reductionism plus the deceptive attractions of the fallacy that conceptual non-reduction is the same as ontological non-reduction.

Thus a genuinely non-reductivist view must be explicit that the emergent properties are genuine natural properties. This Fodor does, for example, by appealing to the role of mental properties in genuine natural laws. That, as O'Connor and Wong (2006) point out leads to a true property dualism, but at the price of overdetermining causes. This, they concede, is softened by 'the usual appeal to an asymmetrical supervenience of the mental on the physical', but, as they go on 'supervenience here will be inexplicable'. (This they contrast with the grounding in causal relations that there is in their brand of emergentism.)

Thus the challenge is no so much overdetermination, but rather overdetermination in the absence of a satisfactorily grounded supervenience. How is it that irreducible higher level (biological, mental) properties can get to supervene on the physical? I have assumed all along that in some way other than they do so supervene; but *why* do they supervene? In the absence of a satisfactory explanation it looks as if alignment of the mental with the physical could be just a fluke. In which case the exclusion problem begins to bite: why isn't it that the causal efficacy of the physical exclude the

mental from having a role? And couldn't the mental and the physical come apart so that they compete, perhaps interfering with one another's causal processes? In which case the causal closure of the physical looks like an accidental feature of the world.

The benefit of the evolutionary emergentism being proposed here is that it both allows the existence of genuinely natural, quasi-fundamental emergent properties and explains their supervenience on the physical, fully fundamental properties. The story of evolutionary emergence in section 2 shows how emergent properties are functional/dispositional properties that have physical realizers. Once a physical system has reached a certain degree and kind of complexity, its (complex) parts will interact so that certain macro-features, functions, are selected for in a manner that is blind to structure. The blindness to structure is what means that emergent properties are irreducible properties. But blindness to structure doesn't mean entirely independent of structure—there has to be some structure supporting the function. And thus we won't have difference in function without difference in structure.¹⁰

Furthermore, we can see why we can appeal to supervenience for emergent properties to avoid the exclusion problem despite the earlier criticism of Loewer's similar appeal. The problem there was that we had no reason for thinking that the higher level universal should always be accompanied by a realizer—why in the nearest world where S is (higher level) M but not (lower level) N, S is N* (some other lower level realizer). Now we do have an explanation for emergent properties, one which will not carry across to non-emergent dispositions (such as fragility). For the nature of such properties is that they are instantiated precisely by the selection of some physical property to realize them. This is the significance of the idea of Vicente and van Gulick that the brain selectively recruits and activates neural structures according to their function. The nearest world in which the higher level, in this case mental, state is present (but not N), is one in which the same selection processes are at work and which thus select some other state (N*) according to its possessing the same functional capacity.

4 Conclusion

The causal completeness of the physical seems to leave little room for genuine causation at higher levels, and in particular within the realm of the special sciences. Whatever causation is to be found there would seem really to be just physical causation, and correspondingly the properties of the special sciences can be causally efficacious only if identical to physical properties or complexes thereof. Thus we can accept causation in the special sciences only at the cost of adhering to reductionism. But reductionism is widely rejected, precisely because there is a failure of type-type identity between higher level and physical properties. At the same time, eliminativism, which concludes that there are no higher level properties at all, is widely resisted. The possibility of *emergent* properties is supposed to square this apparent circle.

The claim of this paper is that the complexity of systems that show selection for functional properties is that answer to what (at least some, more predominant) emergent

¹⁰This is assuming that function can be fixed synchronically, as on the current contribution to fitness view. But if we adopt the diachronic, aetiological view of function, we have to say that we cannot have a difference in function without a difference in *the history of* the structure.

properties are. In such a system it is such functional properties that have causal power, because the development of the system (the interactions of its parts) are sensitive to them in a way that they are not sensitive to the physical causal basis. They are not sensitive to the causal basis, since the causal basis exists precisely to fulfil the role of basis to the function. Had matters been different, the same higher level properties would have been present but with a different causal basis. This is the reverse of the relationship between a causally inert higher level ('abundant') property, such as fragility. In such cases the higher level property is merely a disposition possessed solely in virtue of the object's causal micro-structure, whose presence cannot be explained in terms of the disposition it confers. A fragile snowflake does not have the causal basis for fragility in virtue of its conferring fragility on the snowflake. Nonetheless, emergent properties do have their instantiation explained in terms of their functional character. They are instantiated precisely because a selection process ensures that some physical state fulfilling the appropriate role is instantiated too, thereby ensuring that—and so explaining why—some version of the supervenience thesis holds.

References

- Armstrong, D. M. 1978. *A Theory of Universals. Universals and Scientific Realism Volume II*. Cambridge: Cambridge University Press.
- Batterman, R. 2001. *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*. Oxford: Oxford University Press.
- Clark, A. 1996. *Being There*. Cambridge, Mass.: MIT Press.
- Clark, A. 2001. *Mindware*. Oxford: Oxford University Press.
- Kim, J. 2002a. Précis of *Mind in a Physical World*. *Philosophy and Phenomenological Research* 65: 639–642.
- Kim, J. 2002b. Responses to critics. *Philosophy and Phenomenological Research* 65: 670–679.
- Loewer, B. 2002. Comments on Jaegwon Kim's *Mind in a Physical World*. *Philosophy and Phenomenological Research* 65: 654–661.
- O'Connor, T. and H. Y. Wong 2006. The metaphysics of emergence. *Noûs*.
- Rosenberg, A. 1994. *Instrumental Biology. Or the Disunity of Science*. Chicago: University of Chicago Press.
- van Gulick, R. 1993. Who's in charge here? And who's doing all the work? In Heil and Mele (Eds.), *Mental Causation*. Oxford: Clarendon Press.
- Vicente, A. 2002. How dispositions can be causally relevant. *Erkenntnis* 56: 329–344.
- Vicente, A. 2004. The role of dispositions in explanation. *Theoria* 19: 301–310.
- Walsh, D. 1996. Fitness and function. *British Journal for the Philosophy of Science* 47: 553–574.