EMERGENCE, VARIETIES OF EXPLANATION, AND THE GENERALITY OF LAWS*
(Part II)

J. Schröder

Hanse Institute for Advanced Study
Lehmkuhlenbusch, 4, Delmenhorst, Germany, 27753

The principal aim of this paper is to show that a constraint that C.D. Broad imposed on the acceptability of deductions of macroproperties which would show them to be non-emergent, viz. that they use only general laws of nature, is too strong and should be replaced by the weaker condition that the deductions be non-trivial. First, the relevant notion of generality is made more precise. I propose that a law is general iff it is applicable to a diversity of phenomena relative to what I call "domain constitutive properties". In order to substantiate the claim that Broad's constraint is too strong I analyse three examples of explanations of macroproperties from robotics and the life sciences. All of them are non-trivial explanations and should thereby render the explained properties non-emergent. Finally, I briefly indicate three ways in which an explanation may be non-trivial.

Key words: emergence, laws of nature, reductive explanation, macro properties.

Varieties of explanations of macro properties

In this section I would like to contrast Broad's picture of how the explanation of a macro property would proceed with three examples of explanations which are equally directed at making the emergence of a macro property intelligible, but which reach this goal in a different way. None of these explanations uses composition laws which combine properties of the parts of a system with properties of the whole. Instead the explanations are based exclusively on certain rules or regularities which describe the behaviour of the parts. But, as can be seen from the quote of Broad above, the requirement that in a deduction of a macro property there have to be used only general laws applies to the regularities concerning parts as well as to those which connect the properties of parts with properties of wholes.

The first of these examples is situated in the fields of artificial intelligence and artificial life and tries to explain how adaptive behaviour of a system in its environment comes about. In AI a common strategy to construct (and thereby explain) a system displaying some adaptive behaviour consists in the decomposition of a task in subtasks until very simple subtasks like comparing two digits are reached. The system is then accordingly hierarchically structured in components or modules which are dedicated to the various subtasks. This explanatory strategy is known as homuncularism [10; 18]. For each subtask there is something like a homunculus who carries out the task. Since such hierarchical systems have certain weaknesses like fault intolerance (when a single component breaks down the whole system fails) and dependence on prior analysis of what kinds of events are possible in a certain environment (if something unforeseen turns up the system may fail completely) there have been efforts to circum-

* Статья публикуется в двух номерах; начало опубликовано в № 1 2011 г.

vent these weaknesses by building systems which are not hierarchically structured and functionally decomposable but reach a certain goal via the interaction of components each of which may be directly connected with sensors and effectors and be simultaneously active.

As an example consider a robot's task to follow a wall [25]. The decompositional strategy would try to decompose the whole task in, for example, going to the wall and following it. The subtask of going to the wall could be further decomposed in detecting the wall and moving towards it. The detecting may decompose in sensor reading and sensor interpreting. Alternatively there may be three components which implement the following behaviours:

„1. *Stroll*: Move forward in a more or less straight line motion when there is no obstacle in front. Stop when there is an obstacle. Back up when the obstacle persists.

2. *Avoid*: Turn right or left when there is an obstacle in front of the robot.

3. *Align*: Turn by a small angle when the distance to the object behind is shorter then the distance in front." [Ibid. P. 455].

The components corresponding to these behaviours can all be simultaneously active because the conditions which trigger the behaviours are mutually exclusive so that it will not happen, e.g., that the robot tries to turn right and to move forward at the same time.

It can be „deduced" from these behaviours that the robot will go to a wall (1) (by 1.), then position itself in a right angle to the wall (by 2.) and follow the wall (by 1. and 3.). Given this description of the behavioural repertoire and the conditions under which the behaviours are triggered we can „see" that such a robot displays wall following behaviour without there being a decomposition of the main task in subtasks.

Now, what has all this to do with emergence? From the decompositional point of view there *cannot* be any emergent behaviour because the decomposition of the main task (the candidate for emergent behaviour) guarantees the immediate intelligibility from one level of decomposition to the next. For example, if the subtask of going to the wall is decomposed in detecting the wall and moving in the direction in which the wall has been detected one sees immediately that the subtask will be accomplished because such a movement *constitutes* going to the wall. If there are no equivalent decompositions an impression of necessity (and sufficiency at the same time) imposes itself. If there are equivalent decompositions (which is usually the case in complex and abstract behaviours like saying something) we see at least that the decomposition is sufficient for carrying out the main task.

If an overall behaviour consists of various parts which have to be arranged into a certain sequence decomposition consists simply in analysing this behaviour into those parts. If there is more than one possible analysis every analysis receives its intelligibility in virtue of the fact that in the sequence of subtasks there is a sequence of outputs of these tasks each of which serves as input to the next subtask. This identity of outputs and inputs goes a long way towards explaining our impression that how the macro property or behaviour results from the workings of the parts of a system is just plain. Contrast this understanding with a situation in which there were no such identity, i.e.

in which the inputs a mechanism needed would not be delivered from another mechanism earlier in the sequence, and there would be no explanation. Imagine an arrangement of carburettor, fuel tank, and combustion engine in this order. The carburettor inputs nothing to the fuel tank and the tank inputs liquid fuel into the engine. The engine will not work in this arrangement because it needs another sort of input, fuel in spray form.

When we are successful at breaking down an overall task in subtasks it is of course our previous knowledge and our imagination of the possible arrangement of subtasks which will do the job. And this is the main reason why we don't have the impression that the overall behaviour is emergent. The fact that we have found a decomposition based on our prior knowledge about subtasks and how they combine is enough to convince us that there is no emergence involved. To be sure, when people like engineers go about their business of *looking for* possible decompositions their work may be very difficult. It is not always easy to come up with a solution of the decomposition problem. If you only have a description of the overall behaviour and nothing else, in all the interesting cases it is hardly plain what the decomposition should or might be. But once one is found and you consider it it is not difficult at all to see how the subtasks connect and yield the overall behaviour.

The same does only hold good in a few cases in which there is no hierarchical decomposition. In the example above we are able to understand without difficulty that the wall following behaviour is a necessary consequence from 1., 2., and 3. In the typical case, however, when we only contemplate the rules such understanding is not forthcoming.

Consider another example of the non-decompositional strategy. Suppose you would like to build artificial termites whose task would be to arrange scattered wooden chips into piles. Endow your artificial creatures with mechanisms conforming to the following two rules:

a) If you are not carrying anything and you bump into another wood chip, pick it up.

b) If you are carrying a wood chip and you bump into another wood chip, put down the wood chip you are carrying [21. P. 234].

Why does this work? Can we see from the rules alone that it works? As to the second question I think it is pretty obvious that from these rules we don't easily see that they would yield the desired result. The first rule does not only allow the picking up of a solitary chip but also the removal of a chip which belongs already to a pile. Why should a situation in which the termites start with some thousand solitary chips end up in a situation with a few piles? The answer is that it follows from a) and b) that once the last wood chip is removed from a location there will be never again a wood chip at this location. And this is another way of saying that the number of locations of wood chips is successively reduced and thus the arrangement in piles is guaranteed given enough time.

In this example it is perhaps not necessary to run a computer simulation in order to show that the piles of wood chips are really growing. Once you have hit onto the crucial implication of the rules you can deduce or „see“ that these rules must have the intended effect. There are still other cases, however, where the system property in question does only show up in a computer simulation so that it is clearly invisible from the rules describing the behaviour of the parts of a system. In the last example to be considered it is even harder to see that the emergent property follows from the behaviour of the parts of a system and their interaction without performing such a simulation initially.

There are solitary fish like sharks, moving through the waters on their own, and school fish like sardines which move together with their conspecifics. If you put a fish school into an aquarium which consists of two areas one of which contains food whereas the other does not then, under certain conditions, the following behavior can be observed: The fish school, once it has entered the food area, will remain inside this area. Solitary fish, however, would swim into the food area and out of it and after a while they would enter it again showing a rather homogenous distribution of trajectories across the entire range of the aquarium [22]. This curious behaviour of the fish school can be considered as an emergent property of the school. A property which school fish only display when they are parts of schools but not in isolation. How can this property be explained?

In this example the system is an aggregate entity consisting of individual organisms, viz. a fish school consisting of individual fish. Schooling fish display certain behavioural characteristics or dispositions:

1. If a neighbour's distance exceeds the range of visibility it is not considered any more.
2. If the distance between fish is large but they still can see each other they attract each other.
3. If they are swimming in a school they orient themselves in parallel to their neighbours within the range of the preferred swimming distance.
4. If they get too close to each other they repel each other.
5. If they are feeding they slow down their speed of swimming [3; 6].

In order to run a computer simulation of the trajectories of the school it would be necessary to specify certain values which I have left undetermined in the description of the rules. For example, we would have to determine a value for the range of visibility, another value for the minimal distance below which they repel each other and so on. But when we specify these values we can do the simulations and when we represent the trajectories of the single fish pictorially as lines on a plane we literally see the emergent property, that is, we see that once the lines go into the food area they remain there.

To be sure there is an explanation of this property which does not strictly need the simulation to be carried out. When the school moves from the food area towards the non-food area the first fish leaving the food area increase their speed. This has the effect that they leave the others behind which are still in the food area because of the difference of speed (according to rule 5.). According to rule 2., however, the ones which "break out" will change their swimming direction and thus get back again into the food area. Only if they move in a line perpendicular to the border between the food and the non-food area will there be an escape out of the food area since the only other fish they see are their neighbours to the right and to the left and these neighbours have all the same speed as they themselves. In consequence they will not detect a change of distance and accordingly will not change their direction of movement.

But even if there is such an explanation of the emergent property which does only take into account the rules pertaining to individual behaviour it is an explanation which, in the context of discovery, comes *after* the simulation has been done and which is very unlikely to be hit upon *before* a simulation of the trajectories had been carried through. You have to see that the trajectories of school fish stay mainly in the food area first and then you can go about finding out what, in terms of rules 2. and 5., is responsible for this fact.

What seems to be needed in this case is the iterated application of the rules to the new configuration (position of the fish relative to each other). Only if such iteration is done a huge number of times the emergent property can be finally appreciated. But, and this is the important point, it can only be appreciated if the behaviour of the whole is viewed in a certain way. If we had only information about the trajectories of single fish after 20,000 iterations of the rules and we had this information in a propositional format, e.g. a list of space and time co-ordinates for every member of the school we may still fail to see the emergent property. Only when this information is brought into a pictorial format such that the lists are translated into trajectories the emergent property becomes readily visible. Then it becomes clear that these trajectories lie almost exclusively in that region of the aquarium where the food is located. Of course it is possible to see that the trajectories don't leave the food region when these trajectories are represented in the form of lists. Suppose the food region extends along the y-axis of the aquarium up to point 5. Then, in order to ascertain that the trajectories of all the fish lie inside this region you would have to look up every entry in the list *and*, more importantly, you would have to have *the hypothesis* that they lie in this region beforehand. But the emergent property was supposed to be perceived as a *result* of the process of iteration and not something to be hypothesised. It seems therefore that if the data the iterative process yields are not portrayed in a format adapted to our special cognitive capacities (in this case the visual capacity of pattern recognition) emergent properties will go unnoticed.

Generality of laws versus non-triviality of deductions

At first sight the given examples seem to illustrate only weak emergence, that is, a property which is instantiated only by a certain whole but not by any of its parts. They fall short of the more restricted condition of strong emergence which says that the property of the whole cannot be reductively explained by properties of the parts (2) since in each case the emergent behaviour of the whole could be accounted for by some simple rules pertaining to the behaviour of the parts, either by inspecting the rules themselves or by trying to see what they imply or by running a computer simulation and visualizing the steps of the simulation. When we abstract from the differences between these cases which are of a completely epistemic nature we have an explanation of emergent properties which is non-trivial since it does not use a rule of the form „If the parts of a whole have properties a, b, and c the whole has the properties d, e, and f".

What I would like to emphasise now is that none of the rules used in the explanations qualify as general laws of nature in Broad's sense. Take rule 2 of the first example „Turn right or left when there is an obstacle in front“. This rule is obviously no general law of nature. Why not? It describes a disposition of an object to turn right or left when there is an obstacle in front. As a description of a disposition it could be a general law of nature such as that particles with the same kind of charge (negative or positive) repel each other. But this rule is no law of nature because it does not specify the things for which it is supposed to hold. It might be thought that this problem is not difficult to solve. We could simply stipulate a domain of things. We could, for example, take the domain of all natural and artificial creatures. But then, of course, it is immediately patent that the law would be false because it is falsified by organisms which don't turn left or right when the obstacle is a prey which can easily be captured. Again they don't turn left or right when it is an predator. In this case they move away from it. On the other hand there are artificial creatures, or at least there could be, which move towards an obstacle and just stay there. In any case, it would be no problem to construct such robots. It is only that their behaviour is not very interesting and thus a waste of time and energy. Since natural and artificial creatures would not make up an appropriate domain of validity for our law and since for almost any other natural and artificial domain the same is true we can conclude that there is no domain for which this rule would be a general law. It is valid for robots which have been constructed with the aim that they obey it and this is the reason why it is true of them. But it is not valid for any other domain. The same goes for the rules a) and b) of the second example. Systems for which these rules are true descriptions of their dispositions constitute no natural domain. If, for example termites satisfy it and no other insects it is not a *general* law simply because a larger domain is lacking in which there are different types of organisms each type satisfying the same description. But suppose there were other species for which these rules would be equally valid. Suppose that a variety of other insects behave in a similar way. As long as we don't have a domain constitutive property which determines the domain in which the law is supported to be valid we don't *know* if rules a) and b) are general laws since a law is general only if it holds for every phenomenon of a certain domain. As long as we don't know whether the rule is a general law or not we should be uncertain whether the macroproperty is strongly emergent or not. But as a matter of fact we don't experience any such uncertainty. Instead we feel that we have a perfectly acceptable explanation of the macro property and thus nothing that is strongly emergent.

If natural and artificial termites satisfy the rules they are still no general laws because the laws' generality should be independent of our actions. But once we have understood the workings of certain natural systems we could construct (in principle) an artificial system as a duplicate of the natural one and thereby give the property of generality to a rule which it formerly did not possess. The generality of laws would be trivialised by this procedure and for this reason it cannot depend on our actions.

The third example raises a difficulty, however, not encountered in the other examples. There are school fish and solitary fish. Sardines and thuna are examples of the first, sharks and sword fish of the second kind. The basic rule concerning the behaviour of single school fish, viz. that when one of them sees another one it will approach the other is clearly not only valid for sardines, but also for thuna fish. So it seems that we have a general rule here, a general schoolfish rule. A little reflection, however, shows that this rule has quite a peculiar status, since it is a rule by which the domain of schoolfish is *defined.* Newton's law of gravitation is not of the same character. It does not define the domain for which it is valid. The role of a domain defining property is played by mass. This leads us to postulate a further constraint on general laws: they must not define the domain for which they are valid. Since if they define this domain their truth is a consequence of the definition. They are true of all and only those things they are true of. But „real" general laws should not be true by definition but because the empirical world makes them true, that is, they should not be logically true. If we impose this further constraint the basic rule for schoolfish is not a general law, contrary to first appearances.

In none of the examples, then, there is a general law of nature which is used in the explanation of the macro property. According to Broad's concept of emergence all these macro properties would be emergent in the strong sense since they would not be deducible from properties of the parts and composition rules which are general laws. But this result is certainly counterintuitive because despite our having an explanation of each macro property they should be (according to the generality condition) considered as strongly emergent all the same.

What are we to do with this situation? I think that we can take it to show that the criterion of being a general law is indeed too strong and that it should be replaced by something less restrictive. Why should we draw this moral, why not say that a lot or maybe all of the phenomena with which the sciences of complex systems are concerned are emergent in a strong sense? The crucial consideration here is that what Broad wanted to do with his constraint of general laws was to prevent trivial deductions. The generality of laws was an efficient means to exclude such deductions. With this constraint in place it was not acceptable any more to derive salt's property to dissolve in water from a rule saying „If sodium is combined with chlorine then the resulting compound dissolves in water." But although this constraint does its work very well, it does it too well since it does not only exclude trivial deductions but also excludes deductions which are non-trivial but which don't employ general laws. Since Broad's basic end was to avoid trivial deductions and since his proposed means for achieving this end turns out to be too strong we can simply replace it with Broad's wider constraint that the employed rules must not be trivial. Triviality, in this context, is easily defined. A composition rule is trivial iff it simply connects the properties of the parts with the property of the whole and if this connection is a fact of experience. To be a fact of experience means that only a correlation between a given property of a whole and properties of its parts has been observed and that nothing is known about *how* the properties of the parts conspire to yield the preoperty of the whole. Moreover, in the relevant cases, the properties of the parts are different from those of the whole, e.g. having two electrical poles (H_2O molecules) and being liquid (water). If the property of the whole and of the parts is qualitatively the same then there is no problem, e.g. negatively charged cathode rays which are composed of negatively charged electrons. What Broad had in mind as composition laws which would trivialize deductions were laws connecting different properties as a brute fact, as it were.

Accordingly there are several possibilities for a deduction to be non-trivial (1). Instead of mentioning the observable properties of the parts the composition rule mentions one or more of their theoretical properties. In the case of chemistry, for example, it may mention the number of electrons in various orbitals [cf. 23. P. 440]. Thus a composition rule would be non-trivial if it connects theoretical properties of the parts with other non-observable part properties which are then *conceptually* connected with an observable property of the whole. A case in point is the liquid state of substances. Liquidity is explained, first, by the strength of the forces between the atoms or molecules of a substance. If this strength is small enough the component parts of a substance can move relative to each other. And such motion *is* liquidity. (2) The connection between the observable part properties and the observable macro property is not a fact of experience but a theoretical consequence. For example, when we have a theory which connects the electrical ductility of a piece of metal with the observed behaviour of the metal atoms under certain conditions we are able to predict whether a given substance is a good conductor of electricity if we know how its atoms behave under these conditions (3). The parts of the system have dispositional properties, i.e. their behaviour can be described by rules which together with certain boundary conditions bring about the macro property. To this class of cases our three examples belong. It is important to note that the rules describing the dispositional properties of the parts need not be general laws in the sense discussed.

Every explanation which employs a rule of the form (1), (2), or (3) is a non-trivial explanation and as such renders the macro property non-emergent.

The important lesson that we are taught by these examples is that a condition which otherwise would not have aroused any suspicions does appear in a different light when it is confronted with these cases. In the light of the examples we come to appreciate that Broad's condition is actually too strong. On the other hand, since we understand what kind of service this condition has been intended for we can readily replace it with a wider constraint which still yields the desired result, viz. the exclusion of trivial explanations.

It is somewhat ironic that, exactly as it was the results of the sciences which arrived after Broad had written his book, viz. quantum chemistry and molecular biology, and which turned the properties of chemical compounds and the property of life into non-emergent properties, it is now scientific results (in the form of engineering and computer simulations) which even change the conceptual features of the notion of emergence. As ironic as this may be with respect to Broad I take it to be a good sign. It means that there is progress in philosophy.

NOTES

(1) Provided that there are walls in its environment.

(2) For the concepts of weak and strong emergence see: Stephan A. Emergenz: Von der Unvorhersagbarkeit zur Selbstorganisation [24].

REFERENCES

[1] *Achinstein P.* Law and Explanation. — Oxford: Clarendon Press, 1971.
[2] *Alexander S.* Space, Time and Deity. — London: Macmillan, 1920.
[3] *Aoki I.* A simulation study on the schooling mechanisms in fish // Bulletin of the Japanese Society of Scientific Fisheries, 48. — 1982. — Pp. 1081—1088.
[4] *Bornhofen S., Lattaud C.* "Outlines of Artificial Life: A Brief History of Evolutionary Based Models" // Lecture Notes in Computer Science. — Vol. 3871. Berlin, Heidelberg: Springer. — 2006. — P. 226—237.
[5] *Beckermann A.* Property Physicalism, Reduction and Realization // M. Carrier & P. Machamer (eds.). Mindscapes. Philosophy, Science, and the Mind. — Konstanz/Pittsburgh: Pittsburgh University Press, 1997. — P. 303—321.
[6] *Breckling B., Reuter H.* The Use of Individual Based Models to Study the Interaction of Different Levels of Organization in Ecological Systems // Senckenbergiana maritima 27. — 1996. — P. 195—205.
[7] *Broad C.D.* Mind and its Place in Nature. — London: Routledge & Kegan Paul, 1925.
[8] *Bunge M.* The Mind-Body Problem. A psychobiological Approach. — Oxford: Pergamon, 1980.
[9] *Cariani P.* Emergence and Artificial Life // C. Langton, C. Taylor, J. Farmer, S. Rasmussen (eds.). Artificial Life II. — Addison-Wesley, 1991. — P. 775—97.
[10] *Dennett D.* Brainstorms. — Cambridge, Mass.: MIT Press, 1981.
[11] *Froese T., Ziemke T.* "Enactive artificial intelligence: Investigating the systemic organization of life and mind" // Artificial Intelligence. — 2009. — 173:3—4. — P. 466—500.
[12] *Humphreys P.* Aspects of Emergence // Philosophical Topics. — 1996. — 24. — P. 53—70.
[13] *Humphreys P.* How Properties Emerge // Philosophy of Science. — 1997. — 64. — P. 1—17.
[14] *Hütteman A., Terzidis O.* Emergence in Physics // International Studies in the Philosophy. — 2000. — 14. — P. 267—281.
[15] *Kim J.* "Downward Causation" in Emergentism and Nonreductive Physicalism // A. Beckermann, H. Flohr, J. Kim, (eds.). Emergence or Reduction. — Berlin, New York: de Gruyter, 1992. — P. 119—138.
[16] *Kim J.* Supervenience, Emergence, and Realization in the Philosophy of Mind // M. Carrier and P. Machamer (eds.). Mindscapes: Philosophy, Science, and the Mind. — Konstanz / Pittsburgh: UVK / University of Pittsburgh Press. — 1997. — P. 271—93.
[17] *Lewes G.H.* Problems of Life and Mind. — Vol. 2. — London: Kegan Paul, Trench, Turbner & Co. — 1875.
[18] *Lycan W.* Consciousness. — Cambridge, Mass.: MIT Press, 1987.
[19] *Morgan C. L.* Emergent Evolution. — London: Williams & Norgate, 1923.
[20] *Mill J.St.* A System of Logic. — 4th edn. — London: Parker & Son, 1872.
[21] *Resnick M.* Learning about Life // Artificial Life. — 1994. — 1. — P. 229—42.
[22] *Reuter H., Breckling B.* Selforganization of fish schools: an object oriented model // Ecological Modelling. — 1994. — 75/76. — P. 147—159.
[23] *Schröder J.* Emergence: Non-Deducibility or Downwards Causation? // The Philosophical Quarterly. — 1998. — 48. — P. 433—452.
[24] *Stephan A.* Emergenz: Von der Unvorhersagbarkeit zur Selbstorganisation. — Dresden: Dresden University Press, 1998.
[25] *Steels L.* Towards a Theory of Emergent Functionality // J.-A. Meyer and S.W. Wilson, (eds.). From Animals to Animats. — Cambridge, Mass.: MIT Press, 1991. — P. 451—461.

ЭМЕРДЖЕНТНОСТЬ, МНОГООБРАЗИЕ ОБЪЯСНЕНИЙ И ВСЕОБЩНОСТЬ ЗАКОНОВ (часть II)

Ю. Шрёдер

Hanse Institute for Advanced Study
Лемкуленбуш, 4, Дельменхорст, Германия, 27753

Статья посвящена вопросу об условиях правомерного применения понятия эмерджентности к тем или иным качествам макромира. Главная цель статьи — показать, что введенное Ч.Д. Броудом ограничение на приемлемость дедукций макрокачеств, призванных показать, что макрокачества не являются эмерджентными, является слишком строгим и должно быть заменено более мягким, согласно которому эти дедукции должны быть нетривиальными. Во-первых, следует уточнить применяемое в этом случае понятие всеобщности. По мнению автора, закон является всеобщим, если и только если он применим к многообразию явлений в отношении того, что автор обозначает как «конститутивные свойства данной области» явлений. Для обоснования тезиса о чрезмерной строгости ограничений Броуда в статье анализируется три примера объяснения макрокачеств, взятые из области робототехники и биологических наук. Как показано в статье, все объяснения, приведенные в качестве примеров, нетривиальны, а, следовательно, объясняемые с их помощью качества не являются эмерджентными. В заключение кратко определяются три основные типа нетривиальных объяснений.

Ключевые слова: эмерджентность, законы природы, редуктивное объяснение, макрокачества.