Artificial Mind in a Classical Context

Scott HAGAN and Masayuki HIRAFUJI
Computational Modeling Lab, Dept. of Information Science, N.A.R.C.
3-1-1 Kannondai, Tsukuba, Ibaraki, Japan 305-8666

Abstract

The concept of emergence allows category error to be avoided in the formulation of mental states but must incorporate dynamical criteria to answer questions about when, how and what becomes the property of consciousness. Can we develop enormous complex information processing systems such as our minds ? There must be limit of real time information processing because of restrictions such as speed of light Our objective is to make the limit clear by thought experimental arguments in classical physics. The resulting accounts rely on apprehending global states and classically must be reconciled with two kinds of relativity with respect to external observers. Relativity with respect to the observer's frame can be accommodated by requiring that the emergent property be Lorentz invariant. Relativity with respect to the system boundary can only be reconciled with the ontological character of mental states if these states are intrinsically defined. Local, causal dynamics must therefore be continuously updated as to the location of an intrinsic boundary, a requirement that violates locality constraints. Since there is no pre-existing, ordered space in which mental content can incrementally accrue, consciousness cannot use locality relations to compose perceptual space. Local representations in physical space cannot be made sufficiently complex and distributed representations cannot make the encoded information simultaneously available without violating the locality constraint. The conclusion, that classical science cannot account for all the features of conscious processing, finds convergence in Penrose's recent thesis and might likewise be circumvented in quantum theory.

Keywords: artificial mind, emergence, locality

Introduction

Characteristic features of human mentation suggest that enormous processing advantages might be realized if strategies employed by conscious agents could be artificially reproduced. However theories that directly identify conscious mental features with properties that are physically reducible are guilty of a category error: mental states cannot be identical to physical states unless the two are indiscernible. Theories that scrupulously preserve the distinction between mental and physical, on the other hand, run the risk of leaving mental constructs incapable of causally influencing physical processes. Emergence has arisen, particularly in the context of complex systems, as the prime candidate to circumvent both kinds of error. The purpose of this note will be to establish that no emergent account of conscious mental states is viable within the limits afforded by classical science.

Emergence

The allure of emergence is its ability to accommodate discourse about mental properties without incurring category error. Conflation of the mental and physical ontologies is avoided by postulating a hierarchy of levels, in which "higher-order" properties, corresponding to mental features, arise at some threshold level of complexity.[1] To begin to see how one might then answer questions about what exactly should become the property of consciousness, when the emergent mechanism should come into play and why it should be seen as a necessary consequence of the concomitant conditions, one must first distinguish which varieties of emergent explanation are compatible with a classical scientific characterization.

Any explanation of mental states in the framework of classical science must accommodate the *supervenience* of emergent properties on properties at the level of microphysical dynamics. A class of properties supervenes on a second class of properties whenever two entities, identical with respect to properties in the second class, are of necessity also identical with respect to all properties in the first class. This ensures that two systems, identical at the microphysical level, should yield the same emergent properties, but does not impose the additional, but unnecessary, constraint that each emergent property should be uniquely associated to a particular microphysical configuration. The classical picture further requires that the dynamics at the microphysical level be deterministic. The nature of explanation in the classical model entails that there exist some criteria, possibly inaccessible to empirical observation, on the basis of which the system evolves. This does not in practice imply predictability, nor does it necessarily imply that higher-order descriptions must be reducible to expressions in terms of lower-level processes. A definition of emergence compatible with these restrictions is, for instance, engaged in discussions of chaotic phenomena in complex systems.

With this refined concept of emergence, it becomes possible to specify dynamical criteria on the basis of which answers to questions about what, when and how might be ventured. Models of sensory perception have already begun to speculate about the nature of these criteria, variously identifying perceptual contents with "vector activations"[2], "concentric epicenters"[3] or "attractors."[4] What these accounts have in common is that they all impute mental properties to global states. In fact this is the only option available to an emergent classical theory since any property that can be identified with local states is thereby completely reducible to physical entities.

Emergent properties so constituted are locally implicit as guaranteed by supervenience. Supervenience does not require an explicit representation at the system level, one available and empowered to influence causal dynamics locally. For this reason emergent features are generally characterized as extrinsic. Rendering them explicit requires an act of an observer, external to the system itself, apprehending the whole.

Relativity in Extrinsic Accounts

If mental states are in fact extrinsic, then they must be reconciled with the relative nature of an account given in terms of external observers. The singular character of actual conscious experience must be made compatible with the various accounts afforded from all possible external vantages.

How this might be accomplished can be illustrated in terms of the standard relativity with respect to frame, by a *gedanken* experiment in special relativity. Observer 'A' in the rest frame of a conscious entity will characterize a mental state of that entity as some emergent property of a spatially extended system. Each state is given as a spatially bounded set of events with the same temporal coordinate and from these the observer can create a time series of mental states, the 'thoughts' of the conscious entity according to 'A'. Another observer 'B', in a frame moving with fixed relativistic velocity with respect to the conscious entity, likewise formulates a time series of mental states but, since the definition of simultaneity differs in this frame, will compose each mental state from a different collection of events. The 'thoughts' of the conscious entity according to 'B' need not, *a priori*, match up with those recorded by 'A'. It is clear that, should the accounts differ, only one (at best) can be correct since only one series of mental states actually occurs consciously.

There are two possible resolutions in this case. First, one could require that the time series composed by each possible observer be equivalent to every other. This amounts to imposing a non-trivial constraint, Lorentz invariance, on the formulation of the emergent property as instantiated in terms of physical states. This approach is consistent with both an extrinsic definition of conscious mental states and with the singular quality of actual mental sequences.

Second, one might rule that all accounts from observers outside the rest frame of the conscious entity are incorrect. According to this option, the only correct account that need be given is the one formulated in the rest frame. The determination of content in conscious states is thereby demoted to the level of a perspectival artifact. This might plausibly be maintained only if an extrinsic definition of the emergent state is abandoned. The conscious entity does not, on the view that has been adopted, access and implement one or another of the possible extrinsic accounts. If this were the case, then the account in the rest frame might indeed claim priority over all others. But since the consciousness of the entity has been *defined* on an extrinsic account, there should be no issue of access and no priority of particular cases. What is required then is a shift to an intrinsic definition in which the emergent property is explicitly represented at the system level. Since the conscious system is always trivially in its own rest frame, this gives that frame the desired priority.

To resume, the potential multiplicity of accounts can be accommodated by either (i) imposing Lorentz invariance on an extrinsic formulation or (ii) adopting the perspective of the rest frame in an intrinsic formulation. Relativity is thus resolved in extrinsic theories and repudiated in intrinsic theories.

A second form of relativity to which an emergent characterization of conscious mental states is subject is a relativity with respect to the boundary defining the separation of 'system' and 'environment'. Of course this is, in the context of classical science, a boundary chosen arbitrarily to reflect the observer's convenience concerning which particular system is to be characterized. Since the emergent properties under discussion recognize information and/or relationships available only at the global level, their characterization will be sensitive to a specification of the boundary as it appears in the operative definition of "global." Though certain systems will inevitably be described more conveniently, and some might indeed be more 'interesting', classical phenomena do not in general insist on being circumscribed by particular boundaries. The use of the word 'system' in classical science reflects this element of arbitrariness and the absence of a preferred characterization.

Conscious mental states, in contrast, are not arbitrary. They encompass a certain scope, particular content. Two researchers studying the same conscious entity with different definitions of what exactly constitutes the conscious system might, *a priori*, determine different conscious states. One can, for instance, imagine that the first researcher defines the system boundary to enclose the smallest physical area (or one of the smallest) in which mental properties are manifest while the second researcher chooses a boundary so as to include the first system as a subset (on an extrinsic account, no boundary is to be inherently preferred). The accounts of conscious mental states given by these researchers can presumably be arranged to differ in content but the actual mental state of the studied entity cannot simultaneously satisfy two differing ascriptions of content.

For comparison we might contrast this case with that of another property, 'liquidity', often cited as a paradigm case of emergence. Again we find that systems characterized as 'liquid' are not uniquely defined. The contents of a full glass of water manifest the property, but so too does the bottom half of the water in the glass. Here, however, there is no contradiction in describing 'liquidity' as an emergent property in both systems.

One might try to resolve this issue, and salvage an extrinsic account, by imposing an invariance, analogous to Lorentz invariance in the case of relativity with respect to frame, that would ensure that all possible choices for the boundary would lead to the same mental state. That this is impossible is most trivially evident if we imagine that one researcher delineates the system boundary around the 'brain' of a conscious entity (the 'brain' need not be more critically characterized here) while a second observer draws her own boundary circumscribing the 'brains' of *two* conscious entities.[5] Invariance then implies that the consciousness of the first entity should be equivalent to the combined consciousnesses of both entities and by extension that every consciousness is equivalent to every possible subset of all the consciousnesses in the universe. This, obviously, is not the case.

Locality in Intrinsic Accounts

If an emergent formulation is indeed appropriate, consciousness is distinguished from other emergent phenomena in being independent of whether and how an external observer extracts from the relevant facts about the system and a specification of its boundary, an interpretation that merely renders conscious states. Since no extrinsic account is viable, this independence is possible only if the system intrinsically recognizes that the conditions of satisfaction for a conscious state to emerge are met only with regard to a particular boundary.

Emergent properties are necessarily defined over global states and in order for them to have causal consequences, the global state must inform the local dynamics where the causal mechanisms in classical science are located. Even if the microphysical level is not seen as the sole repository for causal dynamics, an explanation of conscious states compatible with supervenience must allow that there is an interpretation of all causal processes at this level. If the emergent property is to be promoted to intrinsic status, the global state must inform the local dynamics of the current location of its boundary, along with any other information pertinent to its specification or the determination of causal consequences.

In a classical context, locality forbids the fulfillment of this requirement. Locality restricts the speed of propagation of signals over distances, however small. An explicit representation of the boundary cannot possibly be made available to all points in an extended physical system simultaneously. This explains why intrinsically defined systems are nowhere else invoked in classical science.

One might imagine dodging this objection by loosening the requirement of simultaneity and allowing information concerning the global state and its boundary to locally accrue. This strategy requires that physical localities individually assume the burden of storing an explicit representation of all causally empowered features of the global state including a specification of its boundary. This is almost certainly a vast overestimation of the number of degrees of freedom available at individual localities in classical theory, especially once thermal noise is accounted for. Even conservative estimates of the capacity of simple, primitive states far outstrip the potential.[6] There is, moreover, reason to suspect that traditional estimates greatly underestimate the requirements.

Take visual perception as an example. The phenomenal space in which we experience the world visually is ontologically real, not a mere theoretical device, so it will not be acceptable in a fundamental theory to assume a pre-existing structure. Some account of this structure must be given, just as cosmology must account for the structure of physical space-time. Structure in the visual perceptual field is evident in ordering relations like "beside" or "above." Despite the 'spatial' labels, these are not relations in physical space so the mental relations are not explained and maintained simply by the fact that physical entities are correspondingly related.

Part of the task of explaining conscious states is thus the construction of structured mental spaces. This however cannot be accomplished by accrual of information. Assembling mental structure in a piecemeal fashion assumes a pre-existing space in which assembly takes place and which preserves the relations of assembled pieces (regardless of what the 'pieces' are thought to consist in). No locality relations exist in visual perceptual space until the space itself has been constructed so they cannot be used to facilitate the construction. A similarly vicious circle is encountered in cosmology: physical space-time cannot be assembled bit-by-bit by placing points in particular relations to one another since there is no existing concept of locality by which to make sense of these relations. A sensible cosmological theory of space-time structure must necessarily be non-local. A local account is nonsensical; it uses in the explanation the relations that are to be explained.

Mental spaces must similarly be constructed in a non-local way (now intended with respect to mental space). All the necessary relations (of mental primitives) must be provided explicitly and at once. Construction of structure cannot proceed on the basis of partial information. To make use of partial information, there must already be a mental structure to store the relations while more data accumulate. All the data must therefore be simultaneously available in physical space (if mental structure is to derive from relations in physical space).

The classical paradigm provides two options for encoding this information in the physical substrate: local or distributed. A local encoding of all the relations between each mental element and every other cannot be accommodated even under the most absurdly minimal assumptions about the complexity of mental states. And for a distributed representation, locality once again forbids simultaneous access to information that is dispersed over an extended physical region, however small.

Conclusion

We conclude that no emergent account of mental states appears to be possible within the context of classical science. If the emergent property is formulated extrinsically, it is impossible to reconcile the relativity in the definition of the system boundary with the singular nature of experienced conscious states. If it is formulated intrinsically, the locality constraint forbids a spatially extended global system from simultaneously accessing sufficient information to realize a mental state that is intrinsically bounded, complex and ordered.

Evidently the arguments raised with respect to intrinsic formulations are not universal. Intrinsic theories of some complexity might be salvaged if classical systems with a large number of local degrees of freedom could be sufficiently shielded from thermal noise to allow vast numbers of states to be distinguished. There seems little evidence however that the strategies employed biologically make use of entirely local representations. The locality constraint preventing instantiation of physically extended intrinsic representations applies only within the context of classical theory and might be circumvented in theories admitting non-classical elements in a non-trivial way. Penrose's non-computational processes[7] may qualify as would quantum theories if macroscopic quantum states prove, as now appears likely, to be intrinsically bounded. It is in practice problematic to realize such states, and difficult to see what role they could play biologically[8], but if the arguments made here are valid, then such possibilities may hold an essential piece of the puzzle.[9]

---

[1] See for instance Sperry RW (1992) Turnabout on consciousness: a mentalist view. *Journal of Mind and Behavior* 13:259-280.

[2] Churchland PM (1995) The Engine of Reason, the Seat of the Soul: A Philosophical Journey into the Brain. MIT Press: Cambridge, MA.

[3] Greenfield SA (1995) Journey to the Centers of the Mind. W.H. Freeman: New York.

[4] Port RF, Van Gelder T (eds.) (1995) Mind As Motion: Explorations in the Dynamics of Cognition. MIT Press: Cambridge, MA.

[5] Rosenberg GH (1997) A Place for Consciousness: Probing the deep structure of the natural world. Ph.D. dissertation, Indiana University.

[6] Marshall IN (1989) Consciousness and Bose-Einstein condensates. *New Ideas in Psychol.* 7(1):73-83.

[7] Penrose R (1994) Shadows of the Mind. Oxford University Press: Oxford.

[8] Some possibilities are considered in Jibu M, Pribram KH, Yasue K (1996) From conscious experience to memory storage and retrieval: the role of quantum brain dynamics and boson condensation of evanescent photons. *Int. J. Mod. Phys.* B10:1735-1754; Hameroff SR, Penrose R (1996) Conscious events as orchestrated space-time selections. *J. Consciousness Studies* 3:36-53; Jibu M, Hagan S, Pribram K, Hameroff SR, Yasue K (1994) Quantum optical coherence in cytoskeletal microtubules: implications for brain function. *BioSystems* 32:195-209.

[9] See also Stapp HP (1995) Why classical mechanics cannot naturally accommodate consciousness but quantum mechanics can. *Psyche* 2(5) http://psyche.cs.monash.edu.au/v2/psyche-2-05-stapp.html.