Evolutionary Units and Adaptive Intelligence:
Why a brain can not be *produced* by internal evolutionary processes (but could be *composed* of such)

Bruce Edmonds
Centre for Policy Modelling
Manchester Metropolitan University
http://bruce.edmonds.name

Keywords Global brain, intelligence, evolution, computation, embeddedness, unit of evolution, meme, adaptive behaviour

Many people use the term "global brain" to mean any collaborative system (between people or machines) that can do some computation or task that the individual components can't do on their own. This is simply the classic dream of Distributed Artificial Intelligence writ large – lots of bits of computation and thought combine to give a better result. In contrast, in this paper I will take the terms *brain*, *global* and *evolution* literally and discuss the outlook for the actual evolution of a something that could be called global as well as a brain. I contrast the "internally-driven" case where there are lots of co-evolving parts within a system to the "externally-driven" case of a population of systems being evolved in some environment. I argue that these are very different in terms of cause, kind of process and effect.

Brains (systems that implement intelligence)

From an evolutionary perspective the brain is an organ whose function is to help implement adaptive behaviour useful to the organism's propagation. It presumably arose because it conferred selective advantage to the individuals who had it – individuals who were subject to an evolutionary process as a whole person. The brain did not evolve as a result of the internal interactions between neurones, rather the system of internal interactions between neurones (which we external observers distinguish as an organ called 'the brain') evolved as a whole because it conferred selective advantage to the individual.

The main aim of this paper is to point out that a collection of interacting neurones will never become intelligent (as the collection) as a result of purely internal interaction, however sophisticated the neurones, however intricate the interactions, however big the collection and however much time it has to develop. An *external* evolutionary process acting on a population of such collections is required – purely *internal* self-organisation is insufficient. This can framed as an issue about causation in the widest sense: the (external) evolutionary process *caused* the emergence of a brain composed of interacting neurones. The interacting neurones did not (and, I argue, could not) *cause* the emergence of a brain.

I guess some of the confusion comes from an conflation of a brain and a computer, or intelligence and computation – for a computational ability *can* be said to be caused (or can be said to be a direct result of) the its components and their interaction. However a brain is not a computer and a computer is not a brain. A brain can be taught to simulate a computer (given a lot of training by other brains and a high tolerance of tedium), and a computer can simulate aspects of a brain (when programmed to do so by other brains), but this does not mean that they are the same. Analogue processes can be simulated using digital processes and *vice versa* but this does not mean they analogue processes are the same as digital processes.

Another confusion might come from a shallow reading of the work of researchers associated with the artificial life community, for example Kaufman's "The Origins of Order" (Kaufman 1992). One would be forgiven for concluding from this work, that, given the right initial conditions, complex systems might result from simple ones by the deterministic action of rules on each other in a closed 'ecology' of interaction. For example if a set of nodes is randomly connected with an average connectivity of two, the whole network might exhibit a pattern of behaviour which is repetitive only over long time scales. What this shows is that some sorts of system are more capable of sustaining complicated behaviour than others – the existence of such systems might make it easier for an evolutionary process to 'develop' if such a system can be reached as the result of an evolutionary trajectory. None of this sort of work has come up with a non-evolutionary process that is capable of causing the sort of meta-transition (Heylighen 1995) that would create intelligence from non-intelligence, evolution is the *only* known process that can do this.

Different Levels of Evolutionary Process

Another source of confusion seems to come from conflating the various different evolutionary processes that may be taking place.

It seems almost certain that the adaptive behaviour is implemented in the brain using evolutionary mechanisms - e.g. Edelman's neuronal Darwinism (1991) or Dennett's multiple drafts (1992). Thus our brains may *adapt* to its environment (in a lifetime) via an internal process of evolution but only because that is what its been evolved to do by biological evolution.

It also seems likely that some of the benefit of the adaptivity that the brain provides is that it allows sophisticated social coordination which may allow the instigation of a new evolutionary process - that of memes. Thus the existence of brains may allow this new level of memetic evolution to occur.

For an evolutionary process to occur you need units that are: replicated, selected and varied in response to an environment. In the case of humans they evolved a brain to survive and perpetuate themselves in their environment. In the case of memes they evolve as units as those that will be socially perpetuated in the (human) social environment. In the case of neural structures they are copied, reinforced and selected in response to stimulation (which is in turn from the environment via actions and perceptions). Other kinds of units can only evolve adaptive behaviour if they can replicate, vary and be selected out.

The presence of different evolutionary processes at different levels are the source of yet another potential confusion. This train of thought seems to go something like this: *if the brain is (at least partly) implemented using evolutionary processes and we can see that there are (actual and potential) evolutionary processes between brains doesn't that mean that it is possible these could constitute a brain?* The answer is that these internal evolutionary processes can be the building blocks that a brain is implemented on, but these processes will not do the implement*ing* (i.e. the process of building) – an exterior process is necessary for this. That is, an evolutionary process can cause the adaptation of the units it is acting on. Neural Darwinism adapts the synapses between brain cells so they fit in with the brain cells' patterns of excitation. A memetic evolution adapts the memes so they fit in to the culture they are spread in and the brains they are spread into. In biological evolution the organisms are adapted to fit their environment, and brains are part of this adaptation for certain kinds of organism. In no case does a system develop an ability to display adaptive behaviour w.r.t. an environment (i.e. intelligence – Beer 1990) as a result of *internal* evolution, the unit that displays intelligence has to be evolved in a population of others in an environment.

Developing a Culture

Since neither the Internet nor (parts of it) are yet intelligent. Let us look to existing cases of complex interactive (and, at least in the widest sense, evolving) systems – human cultures.

There are (at least) two ways that a culture can be caused to change,

1. as a result of adaptation to some external pressure;
2. as a simple outcome of its members adapting to each other internally.

The tradition in sociology has been that cultures develop internally, that is it composed of internally-drive and generated constructs. The other view is that cultures can adapt societies for the ecological niche they exploit (Reader, 1990) – i.e. they can be understood as externally-driven.

A reasonably clear example of this is the development of academic fields. Academics fields are to different extents constrained by exterior forces (e.g. their subject matter). In some fields the success of a model or idea is more closely determined by the outcomes of observation, experiment or technological utility. In such fields the ideas are adapted to their subject matter to a considerable extent – their content is determined by (or, at least, considerably constrained by) how the world is. In other fields the success of an idea is determined more by how helpful (or simply attractive) it is to other academics in that field. In such fields, the content of the idea is less constrained by an exterior, objective world, but more determined by the nature of the academics and its coherency with the network of accepted ideas in that field.

Of course, all fields are, to some extent, a mixture of these two types but there *are* clear differences in degree. Economics has gone through a stage where the internal tradition has dominated and its connection with its original subject matter is tenuous – it no longer studies the particularities of social phenomena involving exchange of value *per se* but on what ideally 'rational economic men' might do. On the other hand, there have been cases where exterior facts have forced a change despite the prevailing culture. For example, evolution and genetics in the Soviet Union in the last century – there was considerable internal pressure to reject Darwinian evolution for Lamarkian evolution, but it simply did not result in breeding better crops.

This is *not* to say that one style is better than another, it depends upon the field: the point of philosophy can be characterised as producing useful frameworks for thinking about things and thus the attractiveness of its product to an academic could be appropriate; the point of geology is to understand the earth's crust so internal criteria are not so appropriate.

From the outside (e.g. an exterior systems-perspective) the difference can be apparent (for example by the rate to which ideas originating from outside the field are accepted), but what is less clear is how the difference might be from inside. I turn to this next.

The Experience of Being Part of a System

One of the differences in the outcomes between the externally-driven and internally-driven modes of adaptation described above, is that a community of interacting and evolving (as separate units) individuals can become *socially embedded* and improve their own fitness/computational load by utilising the outputs of other agents as 'proxies' for the environment (including the actions of other agents). That is, instead of trying to learn about a complex environment 'from scratch', it may 'short-cut' this process by imitating (or otherwise using) the actions of another individual.

For example, in a stock market, instead of trying to learn to predict future stock prices one might learn to copy a particular trader or even do the opposite of another trader. In turn that trader may be basing its actions on those of still other traders etc. If this web of interaction becomes sufficiently dense the chain of causation can become somewhat disconnected from any market 'fundamentals'. For example if a particularly respected trader starts to sell, many others might copy this, sending the price down, which, in turn, confirms that he was right to sell and that it was sensible to have paid attention to that respected trader. Of course, one can imagine two kinds of market: one externally-driven one where traders look mostly to 'fundamentals' to guide their trades; and another which is internally-driven where speculation and social embedding lead traders to pay more attention to what each other is doing.

I have simulated exactly these sorts social embedding (Edmonds 1999). One of the features I found was that over time in a socially embedded system driven by an internal evolutionary process, the fitness of individuals was relative to the web of internal interaction with other individuals; they are adapted primarily to each other and less to an environment so that their average efficiency marginally dropped. One can think of a maladapted market (dot coms come to mind), where the traders ignore whether the stocks are likely to pay a dividend and effectively conspire to push up prices – the rising market keeps their investors happy, keep themselves rich and secure and is self confirming (for a while).

Whereas if the *whole* system is being evolved as a whole unit (against other whole units) the fitness of each of the individuals in these systems is (at least somewhat) sacrificed to improve the fitness of the whole (as our brains cells are for us). When something happens to make the traders look to the fundamentals of their stocks again (say lots of dot coms going bust and loosing their client's money), the whole market may regain some of its efficiency at the cost of making life much more uncomfortable for the traders.

Another difference is in the source of *meaning* in the system. In a socially embedded system the meaning of interactions is substantially in terms of the tangled history of past interactions, the significance of a communication being in the context of the 'culture' that has developed, whilst in a system that is driven more by exterior constraints the meaning will be grounded in this.

Being a part of an externally-driven system is a very different experience from being part of an interacting community of individually evolving actors. The later is *far* more comfortable to be a part of than the later.

Prospects for (part of) the Internet evolving intelligence

Thus the Internet *as a whole* can not become a Global Brain, however sophisticated the internal interactions, but *parts* of the Internet could. This requires that these *parts* become distinct units capable of reliable replication etc. so that each of these distinct units can evolve. The engine of their evolution would probably be our informational needs as users of the Internet (at least initially).

These units would probably have to evolve to become quite complex entities - well entrenched using their own replicatory processes - before it would be worth the "cost" of supporting brains of their own. To get an idea of the likely scale that would be necessary, remember that humans have about 100,000,000,000,000 cells in their bodies, each of which is a very complex entity in its own right, of which only about 1 in 1,000 are neurones. If the selective advantage that the brain gives is (at least partially) because it allows *societies* to specialise in different environmental niches (Man on Earth), then this implies the necessity of 1,000,000s of humans (i.e. 1000s of societies each composed of 1000s). If we translate one cell into 10,000 Internet nodes this comes to a necessary Internet size of at least 1000,000,000,000,000,000,000,000 nodes. Google currently indexes less than 2000,000,000 nodes, but better estimates of web size are that this is but 10% of the web, so say it is currently 20,000,000,000 nodes. Thus the web would have to grow by roughly a factor of 50,000,000,000,000 before human-type brains will be worth the evolutionary cost. Thus, in the unlikely case of the Internet continuing growing at around the current rate (say doubling every year) this will not occur around the year 2040. Before the Internet has grown by that amount it seems unlikely.

This is probably a low minimum – it is the result of drawing a gene-node analogy, whilst a molecule-node analogy is probably more accurate. The point is that the Internet (or other communication web) will probably *not* be the mechanism for an evolutionary process directly – the nodes and connections won't evolve – but more likely I could be the *substrate* in which structures that can evolve may arise.

Efforts to *design* processes *inside* the Internet to make the *whole network* intelligent will probably fail - what you will get is not a brain but a computer. If a brain is characterised by adaptive intelligence rather than mere computation then this will require the abandonment of a detailed *design* stance and the adoption of a more evolutionary approach (Edmonds 2000). There is no *design* for a brain, it is something that co-evolves along with a set of other systems (in the case of the human brain: the body, human societies, the other organisms that humans interact with and perhaps even language itself).

The difference in terms of experience between being part of an externally evolving system and one in which one is simply adapting to the others in the system, that I have pointed out above can be used as an indication of whether the societies we are part of are themselves evolving entities or just by-products of our interaction (including that part of society expressed on-line). If they do evolve, expect life to be quite uncomfortable, with one's adaptation being driven by the success of that society as a whole – even to the extent of one being killed for that society's sake. If there are no competing societies but, essentially one global society, so that no evolution of that society (in the strict sense) is possible then there will be much less pressure, but only the (relatively) comfortable co-adaptation of its parts.

In terms of the Internet, the signs of an evolutionary process occurring will be parts of the Internet becoming identifiable replicating systems, competing with each other. Before we get brains the Internet would have to get a lot bigger and a lot nastier (complex predation-chains developed). Once we have brains in the replicating systems they might find it in their interest to learn to cooperate so that societies of these brains can inhabit different niches in information space. This is a long way off, and at a scale far beyond us, but we might be able to tell whether it is occurring or not.

References

Beer, R. D. (1990) *Intelligence as Adaptive Behavior*. Academic Press.

Dennett D. (1991). *Consciousness explained*. Penguin Books.

Edelman G. (1992). *Bright air, brilliant fire. On the matter of the mind*. Basic Books.

Edmonds, B. (1999). Capturing Social Embeddedness: a Constructivist Approach. *Adaptive Behavior*, 7:323-348.

Edmonds, B. (2000). The Constructability of Artificial Intelligence (as defined by the Turing Test). *Journal of Logic Language and Information*, 9:419-424.

Giere Ronald, N. (1988). *Explaining science : a cognitive approach*. Chicago ; London, University of Chicago Press.

Heylighen F. & Joslyn C. (1995) Towards a Theory of Metasystem Transitions, *World Futures: the Journal of General Evolution* **45:**1-4.

Kaufmann, S. (1992) *The origins of order : self organization and selection in evolution*. Oxford University Press.

Reader, J. (1990) *Man on Earth*. Penguin Books.