

Empiricism in Artificial Life

Eric Silverman and Seth Bullock

School of Computer Studies, University of Leeds, UK
[eric|seth]@comp.leeds.ac.uk

Abstract

Strong artificial life research is often thought to rely on Alife systems as sources of novel empirical data. It is hoped that by augmenting our observations of natural life, this novel data can help settle empirical questions, and thereby separate fundamental properties of living systems from those aspects that are merely contingent on the idiosyncrasies of terrestrial evolution. Some authors have questioned whether this approach can be pursued soundly in the absence of a prior, agreed-upon definition of life. Here we compare Alife's position to that of more orthodox empirical tools that nevertheless suffer from strong theory-dependence. Drawing on these examples, we consider what kind of justification might be needed to underwrite artificial life as empirical enquiry.

In the title of the first international artificial life conference, held over a decade and a half ago, two streams of Alife research were identified—the *synthesis* of life-like or living systems versus their *simulation*. This distinction was perhaps intended to echo that made between strong and weak artificial intelligence, a division that has been readily adopted by the Alife community. While strong AI or Alife is concerned with building *bona fide* examples of real intelligence or real life, the weak strand of research is concerned with improving our understanding of intelligence and life via the construction of models or replicas of natural systems. While the latter branch of research is understood as an orthodox type of scientific or engineering methodology, the former has often been regarded as more problematic. How can we create genuine instances of intelligence and life without a prior understanding of what constitutes valid membership of either category?

The root of the problem is hinted at by the suggestion of deception, falsity or unreality that can sometimes be detected in the meanings of both *synthetic* and *simulated*. We can clarify this hint by considering two distinct meanings of the term “artificial”.

First, the word artificial can be used to describe a *man-made example of something natural* (hereafter denoted *Artificial*). Artificial light, for example, is typically real light that is manufactured rather than generated naturally by the sun, or forest fires, or bioluminescence of some kind.

By contrast, the word artificial can also be used to describe *something that has been designed to closely resemble something else* (hereafter denoted *Artificial*), perhaps through artifice of some kind. For example, artificial lime might be a flavouring designed to taste like real lime, but which is in fact not the real taste of lime even if it tastes indistinguishable from the real thing. Note that whether or not artificial light or lime can be used to settle empirical questions regarding real light or real lime hinges critically on which definition of artificial is thought to apply.

Which meaning of the word artificial is justified in a particular case depends on what criteria we feel must be met in order for something to count as a true instance of a particular category. In the case of light, we have a physical theory that allows us to lump together sunlight, firelight, torchlight, electric light-bulb light, etc., into one category: real light. Given this, we are in a position to pursue “strong artificial light” research, discovering new ways of generating new kinds of light. Weak artificial light research could investigate ways of producing the *appearance* of light via constructing models of light (which involve no light themselves) or perhaps fake lighting (e.g., ways of painting a room or a picture to suggest the presence of light). In doing so, we might learn about how light works and why it appears to us in the ways that it does.

In the case of lime flavours, we have a less well-specified but widely agreed upon notion that the flavour of real lime fruit cannot be lumped together with some flavouring E555 in a category called “real lime flavour”, no matter how convincingly limey E555 happens to taste, unless perhaps E555 is derived directly and straightforwardly from real limes, or can be proven to be chemically identical to real lime. In this situation, “weak artificial lime” researchers could legitimately concern themselves with producing flavourings that resemble real lime in some way. Through this type of research, we might discover a lot of the chemistry, biology and psychology of what makes things taste like lime. Strong artificial lime research is a little more problematic, however, and we will return to why this is a little later.

In this paper we will consider what kind of framework

could underwrite strong artificial life as an empirical pursuit, since it is through the promise of generating useful empirical data that strong artificial life typically gains its strongest support.

Empirical Alife

In their seminal paper, Alan Newell and Herb Simon (1976) provide an account intended to underwrite artificial intelligence as a kind of empirical enquiry into the nature of intelligence. This account reduces to a pair of working hypotheses:

1. A physical symbol system has the necessary and sufficient means for intelligent action.
2. A suitably programmed computer is an example of a physical symbol system.

The extent to which Newell and Simon were successful is debatable, but it remains the case that no comparable, widely adopted framework has been forthcoming within artificial life. However, the absence of such an account has not prevented Alife practitioners from arguing that their work is a form of empirical enquiry. In particular, it has been suggested that computer simulations of living systems could be used to settle empirical questions concerning life itself.

For example, Mark Bedau (1998) considers a thought experiment proposed by Stephen J. Gould (1989). What would happen, Gould asks, if we could rewind the universe, to a point in time before the advent of life on earth. If we gave the primordial soup a quick stir and then let time run forward again to the year 2004, what would we see on the face of the planet? Would we see creatures much the same as ourselves, and dogs and cats and lice and lichen? Or, as he suggests, is there no reason to suppose that we would not see utterly different life-forms? For Gould, this is a thought experiment about contingency. For Bedau it is an opportunity for Alife simulation to settle an empirical question. By formulating a suitable simulation, one that “manifests” important, fundamental properties of evolution and life, and repeatedly running this simulation, we would be able to observe and record the variety of artificial life-forms that evolve, their regularities and diversity. We would then be able to answer Gould without having to rely on our, rather puny, imaginations.

This type of use for Alife systems has been argued for by a number of authors (Bonabeau & Theraulaz, 1994; Ray, 1994; Taylor & Jefferson, 1994; Miller, 1995) who claim that evolutionary biologists face a problem in that they possess scant evidence with which to reconstruct the evolution of life on earth. Such evidence includes the fossil record, and our limited observations of the species that currently surround us. Compounding this claimed paucity of raw data, the authors point out that any evidence that biologists do possess can only be the result of one evolutionary sequence, since life has (presumably) only evolved once on earth. With

only terrestrial life to draw upon in constructing theories to explain phenomena associated with life, biologists are unable to distinguish between aspects of life which are contingent upon the particular historical development of life on earth, and those aspects which are fundamental to life in general.

For these authors, then, one promise of artificial life is to offer whole new datasets which can be examined alongside the one provided by natural evolution. But this approach to empirical Alife is not without its problems. Consider an anecdote intended to reveal the empirical power of computer simulation, as presented by John Casti (1997) in his book *Would-be worlds: How simulation is changing the frontiers of science*.

Casti describes a particular real sporting event (a Super Bowl game) in which the unfancied American football team happened to beat the favourite. Was the underdog simply lucky, Casti wonders? If the same teams had played again, say 100 times, would the same team have won? Pundits and fans will of course never know for sure since there is only a single data point upon which to base any argument (the actual game as it was played). The data massively underdetermines the hypotheses. But wait! Casti introduces the reader to the existence of an American football computer game in which one can force the computer to control both teams (the user merely watches the game unfold). By re-running the Super Bowl many times, with the same teams and same weather conditions, etc., Casti generates a series of new data points. On the basis of this data, he confirms that, indeed, the underdog had not merely been lucky, since under the highly controlled conditions created by the simulation, it won statistically more often than it lost. This story encapsulates the proposed role for simulation as an empirical tool, filling in missing or incomplete data sets and thereby helping to settle empirical questions. But how valid are these new data points? Can they even be classed as data at all?

Imagine that Casti’s computer game had been released after the Super Bowl that he is re-creating. Could it not be possible, perhaps even highly probable, that the game programmers had noted the outcome of the Super Bowl, and tinkered with the computer program such that it tended to replicate this important result? If that were the case, through his controlled experiment, Casti would have discovered something about the ideas of the computer game programmers, not about the teams “simulated” by the computer game. Even if the computer game had been released *before* the Super Bowl in question had been played, ensuring that the programmers could not have been influenced by its result, the programmers could (must?) have been influenced by other results. How could Casti control for this influence? The programmers themselves may not have understood the idiosyncratic biases that they had perhaps inadvertently included in the game. Surely these biases and suchlike render the simulation moribund, at least as a source of empirical data?

Some authors (e.g., Di Paolo, Noble, & Bullock, 2000) have argued that AI simulations cannot be considered to be sources of empirical data precisely because they are so loaded with the tacit pre-theoretical biases of their creators, coloured by their ideas, and polluted by their opinions. But does philosophy of science not tell us that all observations are theory-loaded in this way? When considering this problem with empirical science, Chalmers reaches the following conclusion:

“... however informed by theory an experiment is, there is a strong sense in which the results of an experiment are determined by the world and not by the theories”, and “we cannot make [the] outcomes conform to our theories.” (Chalmers, 1999, pp.39–40).

Can these claims really be made of computer programs? In this paper we will explore what would have to be true for these constraints to hold in the case of strong artificial life. But first, we will consider two examples of apparently orthodox empirical science in order to discover the ways in which “artificially generated” data take part in regular science.

Trans-Cranial Magnetic Stimulation

Research into brain function often employs patients who have suffered brain damage through strokes or head injuries. Brains are examined to determine which areas are damaged, and associations between these damaged areas and the functional deficits exhibited by the patients are postulated. The technique of trans-cranial magnetic stimulation, known as TCMS or TMS, has allowed psychology researchers to extend the scope of this approach through generating temporary “artificial strokes” in normal, healthy patients.

TMS machinery consists of a set of electrodes that are placed on the outside of the skull. The researcher begins by mapping the major areas of a subject’s brain using an MRI scan, and then proceeds to selectively shut off very small areas of the subject’s brain using magnetic pulses. For example, TMS studies have replicated the effects of certain types of seizure (Fujiki & Steward, 1997), and have examined the effects of stimulation of the occipital cortex in patients with early-onset blindness, finding that sensory areas deprived of input begin to function in other sensory modalities (Kujala, Alho, & Naatanen, 2000). TMS can even trigger anomalous emotional responses; after inhibition of the prefrontal cortex via TMS, visual stimuli that might normally trigger a sad response were much more likely to cause a happy reaction, even laughing (Padberg, 2001).

Such methods provide a way for neuroscientists and psychologists to circumvent the lack of sufficient data from lesion studies. Much of cognitive neuropsychology is restricted in this way, forcing researchers to search through lengthy hospital records and medical journals for patients suffering from appropriate injuries. Even worse, appropriate patients may continue to go unnoticed, as some studies

may require finely differentiated neurological deficits that would not normally be tested for by hospital staff. By using TMS in an attempt to mimic the effects of brain damage, researchers gain the ability to manufacture new case studies, and use this new data to establish or undermine theories regarding the functional architecture of the human brain.

TMS success depends on the precise functioning of the machinery; if subjects undergoing the procedure do not actually experience appropriate inhibition of the brain area under consideration, then the results of a study may be useless. Unfortunately there are few guarantees in TMS research. The machinery inhibits neural activity through bursts of electromagnetism which send the neurons into such a frenzy of activity that normal firing patterns are impossible. This effectively disrupts activity in the area under the pulse, but may not actually mimic the effect of a conventional lesion. Similarly, varying the frequency of the electromagnetic pulse may change the resultant effect on the subject. Finally, the pulse is only intended to affect brain areas which are near to the skull surface; however, the pulse penetrates beyond these areas, and the effect of this excess inhibition is not fully understood. Despite these shortcomings, however, use of TMS to simulate brain dysfunction has become a rapidly growing area of research within psychology and neuroscience.

The data derived from TMS is clearly strongly theory-dependent. Using this data as a way of settling empirical questions regarding the brain function of normal people requires researchers to sign up to a “backstory”. This backstory is an account that lumps TMS data and regular lesion data together as examples of real brain-damage data. Despite the fact that TMS brain damage data is artificial, it will remain admissible if neuroscientists read this artificiality in the sense of Artificial: “a man-made example of something natural”. Neuroscientists would reject TMS if they only regarded TMS brain damage as Artificial: “something that has been designed to closely resemble something else”.

Neuroscience Studies of Rats

Studies of rats are very common within the field of neuroscience, given that rat brains are much less problematic to examine and analyse than human brains. Ideally, researchers would be able to make non-invasive, *in situ* recordings of the neural activity of free-living rats as they go about their normal everyday behaviour. Unfortunately, such recordings are currently beyond the state of the art. In their place, neuroscientists must often rely upon studies of artificially prepared rat brains or portions of rat brain, e.g., to determine the neural pathways that are used during various cognitive functions. A study of GABA-containing neurons in the medial geniculate body of the rat provides a useful example (Peruzzi, Bartlett, Smith, & Oliver, 1997). Rats in this study were anaesthetised and dissected, then slices of the cortex were stimulated directly after preparation in this way. This

particular study aimed to determine the possible varieties of connections within the main auditory pathway of the rat, and by extension the possible structure of the human auditory pathway.

Many neuroscientists would not see any problem with the empirical validity of this type of procedure, and it is certainly the case that one cannot currently identify specific connections between neurons without some sort of similar intervention. However, the behaviour of cortical cells in a preserved culture is certainly not the same as the behaviour that the same cells would exhibit during their normal functioning. The whole brain provides an array of stimulation to the area of cortex in question, and this stimulation is modified by a structured external environment which is itself influenced by rat behaviour. Adaptive behaviour research places a great deal of importance on this notion of embodiment and situatedness, arguing that an organism's coupling to its environment is vital to that organism's cognition and behaviour (Brooks, 1991). If this argument holds, then in some sense neuroscience studies of the kind described above could be accused of generating and recording artificial neurological data. In what way do these data apply to real rat behaviour?

In fact, research in this vein proceeds on the assumption that removal of the rat from the environment may actually increase the experimental validity of the study, since in this way environmental or observer interference can be minimised. The absorption or expulsion of chemicals by individual neurons can be monitored and catalogued with high precision, very accurate readings of neuronal activity can be taken very easily, and experimenter mistakes will likely be more apparent. Additionally, the neuroscience community must hold, perhaps tacitly, that any *artificiality* introduced by their experimental procedures (the extent to which the behaviour of rat neurons changes when the rat is removed from its normal environment, or, in this extreme case, the rat's brain is removed from the rat) is only Artificiality, and hence acceptable.

Discussion

The two examples above demonstrate that relatively orthodox empirical tools are never-the-less theory dependent, and that this dependence need be neither straightforward nor explicitly understood. The grounds upon which the TMS and rat studies described above are considered valid methods of collecting data on real brain function are neither formal nor conclusive, but rather constitute a kind of working hypothesis supported by a tacit framework of assumptions. The fact that empirical tools can be employed in the absence of a strict account of their validity is welcome news for strong artificial life, since a workable definition of what is to count as living and what is not appears to be some way off. But the fact that empirical tools do appear to rely on some agreed-upon "backstory" begs the question: what sort of backstory

might similarly support strong artificial life research?

First and foremost, this backstory must offer a convincing (to the Alife community, at least) account of why artificial life deserves to be understood as a source of Artificial data, rather than merely Artificial data. That is, artificial life must count as man-made life, rather than merely resembling life to some extent.

How might this be achieved? Both TMS and rat neuroscience offer the same suggestion. Propose an empirical procedure that starts with a non-controversial example of the class of systems one wishes to explore (e.g., brains) and prepare it in a manner that makes it amenable to empirical investigation. Then argue that any effects of the preparation procedure (e.g., zapping with electromagnetism; anaesthetising, slicing, and shocking; etc.) are neutral, or benign or can be controlled for. Hence, despite the fact that preparation introduces a gap between the object of enquiry and the subject of experimentation, researchers can still claim that the latter offers a window onto the former, although one that is somewhat *indirect*.

In fact, this is an approach to preparing artificial life already taken within some fields, e.g., AI (Artificial Insemination): start by getting hold of some real male and female gametes from the species you wish to artificially synthesise (uncontroversially, we typically take these gametes to be capable of becoming alive) and artificially bring them together in a way that encourages a new living creature to develop. A similar story underpins other forms of manufactured life, e.g., clones, or mutant lifeforms such as various experimental strains of *Drosophila*.

So far, this approach appears analogous to the manufacture of lime flavour through processing real limes—if this processing is not regarded as somehow debasing or polluting, one might feel justified in using the term "real lime flavour" to describe such a product. Unfortunately, this approach will not satisfy a central goal of strong artificial life: the generation of entirely new datasets. Any artificial life produced in this way will be inherently related to existing life. We will not be able to significantly augment the dataset upon which theoretical biology can draw through this approach.¹ Indeed regular biology regards synthesised life of this type (cloned sheep, mutant flies) as straightforwardly falling within its standard remit.

However, the "preparation" that is typically involved in Alife computer simulation appears to be of an entirely different character. Alife researchers typically begin with a system that appears to be entirely removed from the living systems that they seek to investigate—an unprogrammed computer. The "preparation" that this machine undergoes is

¹An alternative tack also fails for a similar reason. Real lime flavour might be successfully manufactured through building an exact copy of real lime molecules and mixing them together in the right way. But this type of synthesis will not produce new data, it simply reproduces existing data.

some kind of programming, which takes place in a largely idiosyncratic and informal fashion, by contrast with standardised practices in experimental neuroscience. Rather than merely being intended to make the system amenable to controlled observation and recording, this kind of preparation appears to bear the entire burden of “animating” the computer—brining it to life. As such it would be hard to argue that this type of preparation is neutral or benign. Rather, it is substantive, and hence suspicious.

Thus, the gap between the class of living things and the Alife systems built to generate empirical data on them looks pretty wide. Rather than being a window onto life-as-it-could-be, from this perspective artificial life might appear to be quintessentially artificial—an attempt to make one thing look like something else. Can this be remedied?

One example of an alternative route is offered by Newell & Simon (1976). Rather than deriving man-made intelligence from uncontroversial examples of natural intelligence, they commence from a more radical position, arguing that computers meet the necessary and sufficient conditions for intelligence by virtue of being, along with brains, physical symbol systems. Once this is established, any computer can be considered to be only a suitable programming away from being a *bona fide* example of real intelligence. Notice that Newell and Simon do not rely on particular similarities between the behaviour or structure of AI systems and natural intelligences. Rather, they attempt to establish first principles from which to demonstrate a fundamental equivalence between computation and intelligence. On this view, programming a computer is just the same kind of preparation employed in orthodox empirical science—it is a way of subjecting computation/intelligence to controlled, critical examination. How might artificial life follow this lead?

Langton (1992) posits that life is some kind of information ecology—“in living systems, a dynamics of information has gained control over the dynamics of energy, which determines the behavior of most non-living systems” (Langton, 1992, p.41). Perhaps this notion could be elaborated upon to form an appropriate framework for strong artificial life?

1. An information ecology provides the necessary and sufficient conditions for life.
2. A suitably programmed computer is an example of an information ecology.

Under such an account, an Alife simulation would be regarded as an instance of an information ecology, and thus a manifestation of life, rather than merely a facsimile of it. The unprogrammed computer becomes a system that is potentially alive, given the appropriate conditions. Just as Newell and Simon view the computer as simply needing the right type of program in order to realise its potential for intelligent action, strong artificial life proponents might view the computer as similar to an unfertilised egg, needing only

the correct stimulation (i.e., appropriate information) to realise its potential for life. From this perspective, the gap between the object of enquiry and the subject of experimentation not only narrows, but entirely disappears; Alife becomes a means for gathering data on real digital life, rather than a means for producing behaviours similar to biological life. Of course we are a long way from being in possession of a mature account of this kind.

Summary

Here, we have considered the potential for strong artificial life to settle empirical questions regarding the nature of life despite the absence of a unifying theory of life itself. Unfortunately, any “data” drawn from an artificial life study would appear to be strongly theory dependent, i.e., whether or not one considers the behaviour of an artificial life system to constitute Artificial life or Artificial life hinges critically on whether one signs up to an appropriate “backstory” or rationale.

We have seen that less controversial empirical tools also suffer strong theory-dependence, yet are in common use despite the absence of an explicit, formal, conclusive account of their validity. What underwrites the use of these tools is an informal set of tacit assumptions convincing experimentalists that, while the tool may not be a direct window onto the phenomena of interest, any attendant indirectness is not problematic. However, it appears that the kind of frameworks that support the scientific use of these tools will not help support strong artificial life. The validity of TMS and rat neuroscience studies stems from the fact that the starting point for these empirical procedures are uncontroversial examples of the systems they are intended to generate data on. (Although the procedures that are applied could be accused of introducing artificiality, the community that employ these types of technique regard this kind of “preparation” as non-problematic.) As a result, any data generated from these procedures is intimately linked to these uncontroversial examples—it is this link from which the validity of the procedure emanates. If artificial life is intended to generate new datasets that are independent of, or distinct from, life-as-we-know-it, this will not do. Indeed, when “artificial life” is derived from living things, as in cloning, the results are regarded as part of terrestrial biology, not as some distinct kind of life.

In light of this, it appears that strong artificial life will require some theory of life to be in place before it can commence. This theory might follow Newell & Simon (1976) in taking the form of an argument that computation is a kind of life. From this position, it would be possible to generate artificial life that was independent from terrestrial life, yet real in the sense of Artificial. The equation of life and computation is a formidable undertaking, yet artificial life is in possession of a few candidate theories that could act as the seeds of such an account. For instance, Langton’s notion

of living systems as those in which “a dynamics of information has gained control over the dynamics of energy” (Langton, 1991, p.41) may be one such seed. However, such Alife ideas must develop and mature considerably before we are in the enviable position of, say, artificial life researchers, who rely upon a well-founded theory that allows them to generate life in entirely novel ways, yet feel secure in claiming this to be real life, not Artificial life. There is of course no stipulation that such a theory of computation/life need be correct. Ultimately, it will be the artificial life community who will decide whether or not any candidate theory is sufficiently compelling.

Acknowledgements

Thanks to Ezequiel Di Paolo, Jason Noble, an editor of a British neuroscience journal, and members of the Biosystems group for useful conversation.

References

- Bedau, M. A. (1998). Philosophical content and method of artificial life. In Bynum, T. W., & Moor, J. H. (Eds.), *The Digital Phoenix: How Computers are Changing Philosophy*, pp. 135–152. Basil Blackwell, Oxford.
- Bonabeau, E. W., & Theraulaz, G. (1994). Why do we need artificial life? *Artificial Life*, *1*, 303–325.
- Brooks, R. A. (1991). Intelligence without reason. In Myopoulos, J., & Reiter, R. (Eds.), *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence*, pp. 569–595. Morgan Kaufmann, San Mateo, CA.
- Casti, J. L. (1997). *Would-be Worlds: How Simulation is Changing the Frontiers of Science*. Wiley, New York.
- Chalmers, A. F. (1999). *What is this thing called science?* (3rd edition). Open University Press, Buckingham.
- Di Paolo, E. A., Noble, J., & Bullock, S. (2000). Simulation models as opaque thought experiments. In Bedau, M. A., McCaskill, J. S., Packard, N. H., & Rasmussen, S. (Eds.), *Artificial Life VII*, pp. 497–506. MIT Press, Cambridge, MA.
- Fujiki, M., & Steward, O. (1997). High frequency transcranial magnetic stimulation mimics the effects of ECS in upregulating astroglial gene expression in the murine CNS. *Molecular Brain Research*, *44*, 301–308.
- Gould, S. J. (1989). *Wonderful Life: The Burgess Shale and the Nature of History*. W. W. Norton, New York.
- Kujala, T., Alho, K., & Naatanen, R. (2000). Cross-modal reorganization of human cortical functions. *Trends in Neuroscience*, *23*, 115–120.
- Langton, C. G. (1991). Life at the edge of chaos. In Langton, C. G., Taylor, C., Farmer, J. D., & Rasmussen, S. (Eds.), *Artificial Life II — SFI Studies in the Sciences of Complexity*, Vol. X, pp. 41–91. Redwood City, California. Addison-Wesley.
- Langton, C. G. (1992). Life at the edge of chaos. In Langton, C. G., Taylor, C., Farmer, J. D., & Rasmussen, S. (Eds.), *Artificial Life II*, pp. 41–91. Addison-Wesley, Redwood City, CA.
- Miller, G. F. (1995). Artificial life as theoretical biology: How to do real science with computer simulation. Cognitive Science Research Paper 378, School of Cognitive and Computing Sciences, University of Sussex, Brighton, UK.
- Newell, A., & Simon, H. A. (1976). Computer science as empirical enquiry: Symbols and search. *Communications of the Association for Computing Machinery*, *19*, 113–126.
- Padberg, F. (2001). Prefrontal cortex modulation of mood and emotionally induced facial expressions: A transcranial magnetic stimulation study. *Journal of Neuropsychiatry and Clinical Neurosciences*, *13*, 206–212.
- Peruzzi, D., Bartlett, E. L., Smith, P. H., & Oliver, D. L. (1997). A monosynaptic GABAergic input from the inferior colliculus to the medial geniculate body in rat. *Journal of Neuroscience*, *17*, 3766–3777.
- Ray, T. S. (1994). An evolutionary approach to synthetic biology: Zen and the art of creating life. *Artificial Life*, *1*, 179–209.
- Taylor, C. E., & Jefferson, D. R. (1994). Artificial life as a tool for biological inquiry. *Artificial Life*, *1*, 1–13.