

# Chapter 3

## Making the Artificial Real

### 3.1 Overview

Having established in Chaps. 1 and 2 a working understanding of the philosophical underpinnings of computer simulation across disciplines, we turn now to a relatively new field which has created a stir throughout the computer science community to investigate questions of artificiality and its effect upon this type of inquiry. Can a simulation create novel datasets which allow us to discover new things about natural systems, or are simulations based on the natural world destined to be mere facsimiles of the systems that inspire them?

For a possible answer we turn to Artificial Life, a field which jumped to prominence in the late 1980s and early 1990s following the first international conferences on the topic. Proponents of ‘Strong’ Alife argue that this new science provides a means for studying ‘life-as-it-could-be’ through creating digital systems that are nevertheless alive (Langton et al. 1989; Ray 1994).

Of course, such a bold claim invites healthy skepticism, so in this chapter we will investigate the difficulties with strong Alife. By discussing first the nature of artificiality in science and simulation, then investigating related theoretical and methodological frameworks in Artificial Intelligence and more traditional sciences, we attempt to uncover a way in which the strong Alife community might justify their field as a bona fide method for producing novel living systems.

This discussion bears significant importance in the further discussion to come. Providing a theoretical background for any given simulation can greatly impact the modeller’s ability to link a simulation to conventional empirical science, and also serves to illuminate the assumptions inherent in the model and their possible impacts. These issues will be discussed further in the following chapter, in which this first section of the text incorporates broader concerns regarding modelling from the biological community in preparation for creating a framework that incorporates social science in the second section.

## **3.2 Strong vs. Weak Alife and AI**

### ***3.2.1 Strong vs. Weak AI: Creating Intelligence***

The AI community has often been characterised as encompassing two separate strands of research: Strong AI, and Weak AI. Strong AI aims to develop computer programmes or devices which exhibit true intelligence; these machines would be aware and sentient in the same way as a human being. Weak AI, in contrast, aims to develop systems which display a facsimile of intelligence; these researchers do not attempt to create an intelligent being electronically, but instead a digital presence which displays the abilities and advantages of an intelligent being, such as natural language comprehension and flexible problem-solving skills.

### ***3.2.2 Strong vs. Weak Alife: Creating Life?***

The title of the first international artificial life conference immediately identified two streams of Alife research: the synthesis of living systems as opposed to their simulation. Perhaps designed to echo the distinction made between strong and weak artificial intelligence, this division has been readily adapted by the Alife community in the following years. While strong Alife attempts to create real organisms in a digital substrate, the weak strand of Alife instead attempts to emulate certain aspects of life (such as adaptability, evolution, and group behaviours) in order to improve our understanding of natural living systems.

### ***3.2.3 Defining Life and Mind***

Both intelligence and life suffer from inherent difficulties in formalisation; debate has raged in psychology and biology about what factors might constitute intelligent or living beings. AI has attempted to define intelligence in innumerable ways since its inception, resulting in such notable concepts as the Turing Test, in which a computer programme or device which can fool a blind observer into believing that it is human is judged to be genuinely intelligent (Turing 1950).

Meanwhile, prominent philosophers of mind have pointed out the flaws in such arguments, as exemplified in the oft-cited Chinese Room dilemma posed by Searle (1980). In this thought experiment, an individual is locked inside a room with an encyclopaedic volume delineating a set of rules allowing for perfect discourse in Chinese. This individual engages in a dialogue with a native Chinese speaker outside the room merely by following his rulebook. Searle contends that since

the speaker himself does not have any innate knowledge of Chinese, his ability to speak Chinese fluently and convincingly through his rulebook does not prove his intelligence. Counter-arguments to Searle's Chinese Room have attempted to evade this apparent conundrum by claiming that, while the individual inside the room has no comprehension of Chinese, the entire system (encompassing the room, rulebook, and individual) does have an intelligent comprehension of Chinese, thus making the system an intelligent whole; Searle of course offered his own rejoinder to this idea (Searle 1980, 1982), though the controversy continues with connectionists and other philosophers continuing to weigh in with their own analyses (Churchland and Churchland 1990; Harnad 2005).

Similarly life as a phenomenon is perhaps equally difficult to define. While most researchers in relevant fields agree that living systems must be able to reproduce independently and display self-directed behaviour, this definition can fall apart when one is presented with exceptional organisms such as viruses, which display that reproductive component but consist of a bare minimum of materials necessary to for that action. Are such organisms still alive, or are they no more than self-reproducing protein strands? Alternatively, are even those protein strands in some way 'alive'? Researchers and philosophers alike continue to dispute the particulars. Alife researchers' claims that they can produce 'real,' digital life become more problematic in this regard, as under such a nebulous framework for what constitutes life, those researchers have quite a lot of latitude under which to make such statements.

### 3.3 Levels of Artificiality

#### 3.3.1 *The Need for Definitions of Artificiality*

The root of the problem of strong Alife is hinted at by the suggestion of unreality or falsity that can be connoted by the terms thus far used to characterise Alife: synthetic or simulated. A clarification of the type of artificiality under consideration could provide a more coherent picture of the type of systems under examination in this type of inquiry, rather than leaving Alife researchers mired in a morass of ill-defined terminology.

Silverman and Bullock (2004) outline a simple two-part definition of the term 'artificial,' each intended to illuminate the disparity between the natural system and the artificial system under consideration. First, the word artificial can be used to denote a man-made example of something natural (hereafter denoted Artificial<sup>1</sup>). Second, the word can be used to describe something that has been designed to closely resemble something else (hereafter denoted Artificial<sup>2</sup>).

### 3.3.2 *Artificial<sup>1</sup>: Examples and Analysis*

Artificial<sup>1</sup> systems are frequently apparent in everyday reality. For example, artificial light sources produce real light which consists of photons in exactly the same manner as natural light sources, but that light is indeed manufactured rather than being produced by the sun or bioluminescence. A major advantage for the ‘strong artificial light’ researcher is that our current scientific understanding provides that researcher with a physical theory which allows us to combine phenomena such as sunlight, firelight, light-bulb light, and other forms of light into the singular category of real light.

Brian Keeley’s example of artificial flavourings (Keeley 1997) shows the limitations of this definition. While an artificial strawberry flavouring might produce sensations in human tastebuds which are indistinguishable from real strawberry flavouring, this artificially-produced compound (which we shall assume has a different molecular structure from the natural compound) not only originates from a different source than the natural compound, but is also a different compound altogether. In this case, while initially appearing to be a real instance of strawberry flavouring, one can make a convincing argument for the inherent artificiality of the manufactured strawberry flavour.

### 3.3.3 *Artificial<sup>2</sup>: Examples and Analysis*

Artificial<sup>2</sup> systems, those designed to closely resemble something else, are similarly plentiful, but soon show the inherent difficulties of relating natural and artificial systems in this context. Returning to the artificial light example, we could imagine an Artificial<sup>2</sup> system which attempts to investigate the properties of light without producing light itself; perhaps by constructing a computational model of optical apparatus, for example, or developing means for replicating the effects of light upon a room using certain architectural and design mechanisms. In this case, our Artificial<sup>2</sup> system would allow us learn about how light works and why it appears to our senses in the ways that it does, but it would not produce real light as in an Artificial<sup>1</sup> system.

Returning to the case of Keeley’s strawberry flavouring, we can place the manufactured strawberry flavouring more comfortably into the category of Artificial<sup>2</sup>. While the compound is inherently useful in that it may provide a great deal of insight into the chemical and biological factors that produce a sensation of strawberry flavour, the compound itself is demonstrably different from the natural flavouring, and therefore cannot be used as a replacement for studying the natural flavouring.

### 3.3.4 *Keeley’s Relationships Between Entities*

In an effort to clarify these complex relationships between natural and artificial systems, Brian Keeley (1997) describes three fundamental ways in which natural and artificial entities can be related:

... (1) entities can be genetically related, that is, they can share a common origin, (2) entities can be functionally related in that they share properties when described at some level of abstraction, and (3) entities can be compositionally related; that is, they can be made of similar parts constructed in similar ways. (Keeley 1997, p. 3, original emphasis)

This description seems to make the Alife researcher’s position even more intractable. The first category seems highly improbable as a potential relationship between natural systems and Alife, given that natural life and digital life cannot share genetic origins. The third category is more useful in the field of robotics perhaps, in which entities could conceivably be constructed which are compositionally similar, or perhaps even identical, to biological systems. The second category, as Keeley notes, seems most important to Alife simulation; establishing a functional relationship between natural life and Alife seems crucial to the acceptance of Alife as empirical enquiry.

## 3.4 ‘Real’ AI: Embodiment and Real-World Functionality

### 3.4.1 *Rodney Brooks and ‘Intelligence Without Reason’*

Rodney Brooks began a movement in robotics research toward a new methodology for robot construction with his landmark paper ‘Intelligence Without Reason’ (Brooks 1991). He advocated a shift towards a focus on embodied systems, or systems that function directly in a complex environment, as opposed to systems designed to emulate intelligent behaviours on a ‘higher’ level. Further, he posited that such embodiment could produce seemingly intelligent behaviour without high-level control structures at all; mobile robots, for example, could use simple rules to walk which when combined with the complexities of real-world environments may produce remarkably adaptable walking behaviours.

For Brooks and his contemporaries, the environment is not something to be abstracted away in an AI task, but something which must be dealt with directly and efficiently. In a manner somewhat analogous to the Turing Test, AI systems must in a sense ‘prove’ their intelligence not through success in the digital realm but through accomplishments rooted in real-world endeavour.

### ***3.4.2 Real-World Functionality in Vision and Cognitive Research***

While embodiment may appear to be less of a concern in AI research related to vision and cognition, as these behaviours can be separated from the embodied organism more readily, such research is often still rooted in real-world environments. Winograd's well-known SHRDLU program (Winograd 1972) could answer questions addressed in natural language that related to a 'block world' that it was able to manipulate; the program could demonstrate knowledge of the properties of this world and the relationships of the various blocks to one another. While the 'block world' itself was an artificial construct, the success of SHRDLU was based upon its ability to interact with the human experimenter about its knowledge of that virtual world, rather than just its ability to manipulate the virtual blocks and function within that world.

Computer vision researchers follow a similar pattern to the natural-language community, focusing on systems which can display a marked degree of competence while encountering realistic visual stimuli. Object recognition is a frequent example of such a problem, testing a system's ability to perceive and recognize images culled from real-world stimuli, such as recognising moving objects in team-sports footage (Bennett et al. 2004). Similarly, the current popularity of CCTV systems has led to great interest in cognitive systems which can analyse the wealth of footage provided from many cameras simultaneously (Dee and Hogg 2006). In the extreme, modern humanoid roboticists must integrate ideas from multiple disciplines related to human behaviour and physiology as well as AI in order to make these constructions viable in a real-world environment (Brooks et al. 1999).

Within the AI research community, the enduring legacy of the Turing Test has created a research environment in which real-world performance must be the end goal; merely understanding basic problems or dealing exclusively in idealised versions of real-world situations are not sufficient to prove a system's capacity for intelligent behaviour. The question of artificiality is less important than that of practicality; as in engineering disciplines, a functional system is the end goal.

### ***3.4.3 The Differing Goals of AI and Alife: Real-World Constraints***

Clearly, real-world constraints are vital to the success of most research endeavours in AI. Intelligent systems, Strong or Weak, must be able to distinguish and respond to physical, visual, or linguistic stimuli, among others, in a manner sufficient to provide a useful real-world response. Without this, such systems would be markedly inferior next to the notable processing abilities of even the most rudimentary mammalian brain.

For Alife, however, the landscape is far more muddled. Most simulations take place in an idealised virtual context, with a minimum of complex and interacting factors, in the hopes of isolating or displaying certain key properties in a clear fashion. Given the disparity between the biological and digital substrates, and the difficulties in defining life itself, the relationship between the idealised virtual context of the simulated organisms and any comparable biological organism seems quite wide.

In this sense, artificial intelligence has a great advantage, as 'human' intelligence and reasoning is a property of a certain subset of living beings, but can be viewed as a property apart from the biological nature of those living beings to a degree. Artificial life, by its very nature, attempts to investigate properties which depend upon that living substrate in a much more direct fashion.

### **3.5 'Real' Alife: Langton and the Information Ecology**

#### ***3.5.1 Early Alife Work and Justifications for Research***

The beginnings of Alife research stemmed from a number of examinations into the properties of life using a series of relatively recent computational methods. Genetic algorithms, which allow programs to follow a process of evolution in line with a defined 'fitness function' to produce more capable programs, were applied to digital organisms which attempted to compete and reproduce rather than simply solve a task (Ray 1994). Similarly, cellular automata displayed remarkable complexity despite being derived from very simple rules; such a concept of 'emergent behaviour' came to underwrite much of Alife in the years to come. Creating simulations which display life-like behaviours using only simple rule-sets seemed a powerful metaphor for the complexity of life deriving from the interactions of genes and proteins.

#### ***3.5.2 Ray and Langton: Creating Digital Life?***

Ray (1994) and Langton (1992) were early proponents of the Strong Alife view. Ray contended that his Tierra simulation, in which small programs competed for memory space in a virtual CPU, displayed an incredible array of varied 'species' of digital organisms. He further posited that such simulations might hail the beginnings of a new field of digital biology, in which the study of digital organisms may teach researchers about new properties of life that may be difficult to study in a natural context; he argued that his digital biosphere was performing fundamentally the same functions as the natural biosphere:

Organic life is viewed as utilising energy, mostly derived from the Sun, to organize matter. By analogy, digital life can be viewed as using CPU (central processing unit) time, to organize memory. Organic life evolves through natural selection as individuals compete for resources (light, food, space, etc.) such that genotypes which leave the most descendants increase in frequency. Digital life evolves through the same process, as replicating algorithms compete for CPU time and memory space, and organisms evolve strategies to exploit one another. (Ray 1996, p. 373-4)

For Ray, then, an environment providing limited resources as a mechanism for driving natural selection and an open-ended evolutionary process is sufficient to produce ‘increasing diversity and complexity in a parallel to the Cambrian explosion.’ He goes on to describe the potential utility of such artificial worlds for a new variety of synthetic biology, comparing these new digital forms of ‘real’ artificial life to established biological life.

Langton (1992), in his investigation of cellular automata, takes this idea one step further, describing how ‘hardware’ computer systems may be designed to achieve the same dynamical behaviours as biological ‘wetware.’ He posits that properly organised synthetic systems can provide these same seemingly unattainable properties (such as life and intelligence), given that each of these types of systems can exhibit similar dynamics:

... if it is properly understood that hardness, wetness, or gaseousness are properties of the organization of matter, rather than properties of the matter itself, then it is only a matter of organization to turn ‘hardware’ into ‘wetware’ and, ultimately, for ‘hardware’ to achieve everything that has been achieved by wetware, and more. (Langton 1992, p. 84)

For Langton, life is a dynamical system which strives to avoid stagnation, proceeding in a series of phase transitions from one higher-level evolved state to the next. This property can be investigated and replicated in computational form, which in his view seems to provide sufficient potential for hardware to develop identical properties to wetware in the appropriate conditions (such as an appropriately-designed cellular automata space).

### 3.5.3 *Langton’s Information Ecology*

Langton (1992) and Langton et al. (1989) attempted to justify his views regarding artificial life by proposing a new definition for biological life. Given that natural life depends upon the exchange and modification of genetic information through natural selection, Langton suggests that these dynamics of information exchange are in fact the essential components of life:

... 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)

Thus, the comparatively simple thermodynamically regulated behaviours of non-living systems give way to living systems regulated by the dynamics of gene exchange. From this premise, Langton proposes that if this ‘information ecology’

were accepted as the defining conditions for life, then a computer simulation could conceivably create an information ecology displaying the same properties.

## 3.6 Toward a Framework for Empirical Alife

### 3.6.1 *A Framework for Empirical Science in AI*

If we seek the construction of a theoretical framework to underwrite empirical exploration in Alife, we can gather inspiration from Newell and Simon's seminal lecture (Newell and Simon 1976). The authors sought to establish AI as a potential means for the empirical examination of intelligence and its origins. They argue that computer science is fundamentally an empirical pursuit:

Computer science is an empirical discipline... Each new machine that is built is an experiment. Actually constructing the machine poses a question to nature; and we listen for the answer by observing the machine in operation and analyzing it by all analytical and measurement means available. (Newell and Simon 1976, p. 114)

However, the argument that computer science is fundamentally empirical due to its experimental interactions with replicable, physical systems is not sufficient to claim that an AI can be used as a means to study intelligence empirically. Newell and Simon address this by proposing a definition of a physical symbol system, or 'a machine that produces through time an evolving collection of symbol structures' (Newell and Simon 1976, p. 116). The details of the definition and its full import are beyond the scope of this text, but in essence, the authors argue that physical symbol systems are capable of exhibiting 'general intelligent action' (Newell and Simon 1976, p. 116), and further, that studying any generally intelligent system will prove that it is, in fact, a physical symbol system.

Newell and Simon then suggest that physical symbol systems can, by definition, be replicated by a universal computer. This leads us to their famous Physical Symbol System Hypothesis – or PSS Hypothesis – which we can summarise as follows:

1. 'A Physical Symbol System has the necessary and sufficient means for general intelligent action.' (Newell and Simon 1976, p. 116)
2. A computer is capable of replicating a Physical Symbol System.

Thus, by establishing general intelligence as a process of manipulating symbols and symbol expressions, and that computers are capable of replicating and performing identical functions – and indeed are quite good at doing so – Newell and Simon present AI as an empirical study of real, physical systems capable of intelligence. AI researchers are not merely manipulating software for the sake of curiosity, but are developing *real examples* of intelligent systems following the same principles as biological intelligence.

### 3.6.2 *Newell and Simon Lead the Way*

Newell and Simon's PSS Hypothesis (Newell and Simon 1976) largely succeeded in providing a framework for AI researchers at the time, and one that continues to be referenced and used today. Such a framework is notably absent in Alife, however, given the difficulties in both defining Alife and in specifying a unified field of Alife, which necessarily spans quite a few methodologies and theoretical backgrounds. With Langton's extension of the fledgling field of Alife into the study of 'life-as-it-could-be,' a more unified theoretical approach seems vital to understanding the relationship between natural and artificial life:

Artificial life is the study of artificial systems that exhibit behavior characteristic of natural living systems. It is the quest to explain life in any of its possible manifestations, without restriction to the particular examples that have evolved on Earth. This includes biological and chemical experiments, computer simulations, and purely theoretical endeavors. Processes occurring on molecular, social and evolutionary scales are subject to investigation. The ultimate goal is to extract the logical form of living systems. (Langton, announcement of Artificial Life: First International Conference on the Simulation and Synthesis of Living Systems)

By likening Alife to the study of the very nature of living systems, Langton appeals to the apparent flexibility and power of computer simulations. The simulation designer has the opportunity to create artificial worlds that run orders of magnitude faster than our own and watch thousands of generations of evolution pass in a short space of time, and thus seems to provide an unprecedented opportunity to observe the grand machinery of life in a manner that is impossible in traditional biology.

This, in turn, leads to an enticing prospect: can such broad-stroke simulations be used to answer pressing empirical questions about natural living systems? Bedau (1998), for example, sees a role for Alife in a thought experiment originally proposed by Gould (1989). Gould asks what might happen if we were able to rewind evolutionary history to a point preceding the first developments of terrestrial life. If we changed some of those initial conditions, perhaps merely by interfering slightly with the 'primordial soup' of self-replicating molecules, what would we see upon returning to our own time? Gould suggests that while we might very well see organisms much the same as ourselves, there is no reason to assume that this would be the case; we may resume life in our usual time frame to discover that evolutionary history has completely rearranged itself as a result of these manipulations.

For Bedau, this thought experiment presents an opening for Alife to settle a question fundamentally closed to traditional biology. By constructing a suitable simulation which replicates the most important elements of biological life and the evolutionary process, and running through numerous simulations based on variously-perturbed primordial soups, we could observe the resultant artificial organisms and see for ourselves the level of diversity which results.

Other authors have proposed Alife simulations that fall along similar lines (Bonabeau and Theraulaz 1994; Ray 1994; Miller 1995), noting that evolutionary biologists are burdened with a paucity of evidence with which to reconstruct the

evolutionary course of life on Earth. The fossil record is notoriously incomplete, and our vanishingly small time on Earth has allowed precious few opportunities to observe even the species which exist around us today. In addition, as Gould's thought experiment highlights, we only have the opportunity to observe those organisms which have evolved on Earth, leaving us entirely uncertain of which properties we observe in that life are particular to life on Earth, and which are particular to life in any form.

Quite obviously such an experiment could be potentially revolutionary, and yet the methodological problems are vast despite the inherent flexibility of computer simulation. How could we possibly confirm that these artificial organisms are Artificial<sup>1</sup> rather than Artificial<sup>2</sup>? Given our lack of a definition for biological life, determining whether a digital organism is alive seems a matter of guesswork at best. Further, how detailed must the simulation be to be considered a reasonable facsimile of real-world evolutionary dynamics? Should they select on the level of individuals, or genes, or perhaps even artificial molecules of some sort? If by some measure we determine the simulation to be Artificial<sup>2</sup> rather than Artificial<sup>1</sup>, can we truly claim that this simulation in any way represents the development of natural living systems, or is it merely a fanciful exploration of poorly-understood evolutionary dynamics?

### 3.6.3 *Theory-Dependence in Empirical Science*

Beyond these methodological considerations, some troubling philosophical issues appear when considering such large-scale applications of the Alife approach. Some have argued that Alife simulations should not, and in fact cannot, be considered useful sources of empirical data given that they are loaded with inherent biases from their creators (Di Paolo et al. 2000). Simulations must necessarily adopt varying levels of abstraction in order to produce simulations which are both computable and analysable; after all, simulations which replicate the complexities of real biology in exacting detail would gain the experimenter very little time-saving and eliminate one of the major benefits of computational modelling. These abstractions are tacitly influenced by the experimenter's own biases, relying upon the simulation creator's perception of which aspects of the biological system can be simplified or even removed from the simulation entirely. Similarly, the parameter settings for that simulation will come with their own biases, resulting in simulations which could produce vastly different results depending on the theoretical leanings of the programmer.

While one might claim that conventional empirical science suffers from very similar shortfalls, Chalmers points out that biases within such fields must necessarily end when the results of an experiment contradict those biases:

... 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 theories... we cannot make [the] outcomes conform to our theories. (Chalmers 1999, p. 39–40)

In the case of computer simulations this conclusion seems more difficult. After all, we are in essence adding an additional layer – in addition to the two layers of the physical world and the theory that describes it, we now have a model layer as well, which approximates the world as described by our theory. Clearly this adds additional complications: when not only the experiment but the very world in which that experiment takes place are designed by that biased experimenter, how can one be sure that pre-existing theoretical biases have not completely contaminated the simulation in question?

To return to our central bird migration example, one can imagine various ways in which our migration researcher could introduce theoretical bias into the simulation. That researcher's ideas regarding the progression of bird migration, the importance of individual behaviours or environmental factors in migration, and other pre-existing theoretical frameworks in use by the researcher will inform the assumptions made during construction of the model. In such a scenario, removing such biases is difficult for the modeller, as on some level the model requires certain assumptions to function; the researcher needs to develop a theoretical framework in which this artificial data remains usable despite these issues.

### ***3.6.4 Artificial Data in Empirical Science***

An examination of more orthodox methods in empirical science may shed some light on how some methodologies use artificially-generated data to answer empirical questions. While these methods are more based in the natural world than an Alife simulation, they still rely upon creating a sort of artificial world in which to examine specific elements of a process or phenomenon.

Despite the artificiality of this generated data, such disciplines have achieved a high degree of acceptance amongst the research community. A brief look at some examples of the use of such data in empirical research will provide some insight into possible means for using artificial data derived from simulation.

#### **3.6.4.1 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 by generating temporary artificial strokes in otherwise healthy individuals.

TMS machinery produces this effect through a set of electrodes which are placed on the outside of the subject's skull. The researcher first maps the subject's brain using an MRI scan, and then begins using the TMS apparatus to fire brief

magnetic pulses at the surface of the brain. These pulses in effect overstimulate the affected neurons, causing small areas of the brain to ‘short-circuit,’ which replicates the effects of certain types of brain injury. TMS researchers have managed to replicate the effects of certain types of seizure (Fujiki and Steward 1997), and have examined the effects of stimulation of the occipital cortex in patients with early-onset blindness (Kujala et al. 2000). TMS has even produced anomalous emotional responses; after inhibition of the prefrontal cortex via TMS, visual stimuli that might normally trigger a negative response were much more likely to cause a positive response (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 suffers from this paucity of raw data, forcing researchers to spend inordinate amounts of time searching through lengthy hospital records and medical journals for patients suffering from potentially appropriate brain injuries. Even when such subjects are found, normal hospital testing may not have revealed the true nature of the subject’s injury, meaning that potentially valuable subjects go unnoticed as some finely differentiated neurological deficits are unlikely to be discovered by hospital staff. Assuming that all of these mitigating factors are bypassed, the researcher is still faced with a vanishingly small subject pool which may adversely affect the generalisability of their conclusions.

TMS thus seems a remarkable innovation, one which suddenly widens the potential pool of subjects for cognitive neuropsychology to such a degree that any human being may become a viable subject regardless of the presence or lack of brain injury. However, there are a number of shortcomings to TMS which could cause one to question the validity of associated results. The inhibition of neural activity produced by TMS causes neurons to explode into such activity that normal firing patterns become impossible; while this certainly effectively disrupts activity in the brain area underneath the pulse, TMS is not necessarily capable of replicating the many and varied ways in which brain injuries may damage the neural tissue. Similarly, the electromagnetic pulse used to cause this inhibition of activity is only intended to affect brain areas which are near to the skull surface; the pulse does not penetrate beyond those areas, but the effects of this excess inhibition are not fully understood. Despite these shortcomings, and the numerous areas in which standard examinations of brain-injury patients are superior, TMS has continued to become a rapidly-growing area of research in cognitive neuropsychology.

The data derived from TMS is strongly theory-dependent in nature, in a similar fashion to computer simulation. In order to use this data as a reasonable method of answering empirical questions requires that TMS researchers adopt a theoretical ‘backstory.’ This backstory is a framework that categorises TMS data and lesion data together as examples of ‘real’ brain-damage data. Despite the fact that TMS brain damage is generated artificially, it is deemed admissible as ‘real’ data as long as neuroscientists consider this data Artificial<sup>1</sup> rather than Artificial<sup>2</sup> in classification.

### 3.6.4.2 Neuroscience Studies of Rats

Studies of rats have long been common within the field of neuroscience, given the complex methodological and ethical difficulties involved in studying the human brain. Though such studies allow much greater flexibility due to their more uncontroversial participants, certain fundamental limitations still prohibit certain techniques; while ideally, researchers would prefer to take non-invasive, in situ recordings of the neural activity of normal, free-living rats, such recordings are well beyond the current state of the art.

Instead, neuroscientists must settle for studies of artificially prepared rat brains or portions thereof; such a technique would be useful if the researcher wishes to determine the activity of specific neural pathways given a predetermined stimulus, for example. One study of the GABA-containing neurons within the medial geniculate body of the rat brain required that the rats in question were anaesthetised and dissected, with slices of the brain then prepared and stimulated directly (Peruzzi et al. 1997). Through this preparatory process, the researchers hoped to determine the arrangement of neural connections within the rat's main auditory pathway, and by extension gain some insight into the possible arrangement of the analogous pathway in humans.

Given how commonplace such techniques have become within modern neuroscience, most researchers in that community would not question the empirical validity of this type of procedure; further, the inherent difficulties involved in attempting to identify such neural pathways within a living brain makes such procedures the only known viable way to take such measurements. However, one might argue that the behaviour of cortical cells in such a preserved culture, entirely separate from the original brain, would certainly differ significantly from the behaviour of those same cells when functioning in their usual context. The entire brain would provide various types of stimulus to the area of cortex under study, with these stimuli in turn varying in response to both the external environment and the rat's own cognitive behaviour.

With modern robotics and studies of adaptive behaviour focusing so strongly upon such notions of embodiment and situatedness, the prevailing wisdom within those fields holds that an organism's coupling to its external environment is fundamental to that organism's cognition and behaviour (Brooks 1991). In that case, neuroscience studies of this sort might very well be accused of generating and recording 'artificial' neural data, producing datasets that differ fundamentally from the real behaviour of such neurons. How then do neuroscientists apply such data to the behaviour of real rats, and in turn to humans and other higher mammals?

In fact, research of this type proceeds on the assumption that such isolation of these cortical slices from normal cognitive activity actually increases the experimental validity of the study. By removing the neurons from their natural environment, the effects of external stimuli can be eliminated, meaning that the experimenters have total control over the stimulus provided to those neurons. In addition, the chemical and electrical responses of each neuron to those precise stimuli can be measured precisely, and the influence of experimenter error can be minimised.

With respect to the artificiality introduced into the study by using such means, neuroscience as a whole appears to have reached a consensus. Though removing these cortical slices from the rat is likely to fundamentally change the behaviour of those neurons as compared to their normal activation patterns, the individual behaviour of each neuron can be measured with much greater precision in this way; thus, the cortical slicing procedure can be viewed as Artificial<sup>1</sup>, as the individual neurons are behaving precisely as they should, despite the overall change in the behaviour of that cortical slice when isolated from its natural context. Given the (perhaps tacit) agreement that these methods are Artificial<sup>1</sup> in nature, the data from such studies can be agreed to consist of ‘real’ neuroscience data.

### 3.6.5 *Artificial Data and the ‘Backstory’*

The two examples above drawn from empirical science demonstrate that even commonplace tools from relatively orthodox fields are nevertheless theory-dependent in their application, and further that such dependence is not necessarily immediately obvious or straightforward. The tacit assumptions underlying the use of TMS and rat studies in neuroscience are not formalised, but they do provide a working hypothesis which allows for the use of data derived from these methods as ‘real’ empirical data. The lack of a conclusive framework showing the validity of these techniques does not exclude them from use in the research community.

This is of course welcome news for the strong Alife community, given that a tangible and formal definition of what constitutes a living system is quite far from being completed. However, strong Alife also lacks this tacit ‘backstory’ that is evident in the described empirical methods; without this backstory those methods may fall afoul of the Artificial<sup>1</sup>/Artificial<sup>2</sup> distinction. With this in mind, how might the Alife community describe a similar backstory which underwrites this research methodology as a valid method for producing new empirical datasets?

The examples of TMS and rat neuroscience offer one possibility. In each of those cases, the investigative procedure begins with an uncontroversial example of the class of system under investigation (i.e., in the case of TMS, a human brain). This system is then prepared in some way in order to become more amenable to a particular brand of investigation; rat neuroscientists, by slicing and treating the cortical tissue, allow themselves much greater access to the functioning of individual neurons. In order to justify these modifications to the original system, these preparatory procedures must be seen as neutral in that they will not distort the resulting data to such a degree that the data becomes worthless. Indeed, in both TMS and rat neuroscience, the research community might argue that these preparatory procedures actually reduce some fairly significant limitations placed upon them by other empirical methodologies.

At first blush such a theoretical framework seems reasonable for Alife. After all, other forms of ‘artificial life’ such as clones or recombinant bacteria begin with such an uncontroversial living system and ‘prepares’ it while still producing

a result universally accepted to be another living system. However, this does fall substantially short of one of the central goals of strong artificial life, as these systems certainly produce augmented datasets regarding the living systems involved, but they do not generate entirely novel datasets. While recombinant bacteria and mutated *Drosophila* may illuminate elements of those particular species that we may have been unable to uncover otherwise, the investigation of ‘life-as-it-could-be’ remains untouched by these efforts.

In addition to these shortcomings, the preparatory procedures involved in a standard Alife simulation are quite far removed from those we see in standard biology (or the neuroscience examples discussed earlier). These simulated systems exist entirely in the digital realm, completely removed from the biological substrate of living systems; though many of these simulated systems may be based upon the form or behaviour of natural living systems, those systems remain separate from their simulated counterparts.

Further, this computer simulation is ‘prepared’ in this digital substrate through a process of programming to produce this artificial system. Programming is inherently a highly variable process, the practice of which differs enormously from one practitioner to the next, in contrast to the highly standardised procedures of neuroscience and biology. The result of these preparations is to make the system amenable to creating life, rather than simply taking a previously-existing system and making it more amenable to a particular type of empirical observation. While characterising this extensive preparatory process as benign, as in neuroscience and biology, would be immensely appealing to the Alife community, the argument that preparing a computer and somehow creating life upon its digital substrate is a benign preparatory procedure is a difficult one to make.

### ***3.6.6 Silverman and Bullock’s Framework: A PSS Hypothesis for Life***

Thus, while we have seen that even orthodox empirical science can be considered strongly theory-dependent, the gap between natural living systems and Alife systems remains intimidatingly wide. From this perspective, characterising Alife as a useful window onto ‘life-as-it-could-be’ is far from easy; in fact, Alife appears dangerously close to being a quintessentially Artificial<sup>2</sup> enterprise.

Newell and Simon’s seminal paper regarding the Physical Symbol System Hypothesis offers a possible answer to this dilemma (Newell and Simon 1976). Faced with a similar separation between the real system of interest (the intelligence displayed by the human brain) and their own attempt to replicate it digitally, the PSS Hypothesis offered a means of justifying their field. By establishing a framework under which their computers offer the ‘necessary and sufficient means’ for intelligent action, Newell and Simon also establish that any computer is only a short (albeit immensely complicated) step away from becoming an example of real, Artificial<sup>1</sup> intelligence.

The PSS Hypothesis also escapes from a potential theoretical conundrum by avoiding discussion of any perceived similarities between the behaviour of AI systems and natural intelligence. Instead Newell and Simon attempt to provide a base-level equivalence between computation and intelligence, arguing that the fundamental symbol-processing abilities displayed by natural intelligence are replicable in any symbol-processing system of sufficient capability. This neatly avoids the shaky ground of equating AI systems to human brains by instead equating intelligence to a form of computation; in this context, the idea that computers can produce a form of intelligence follows naturally.

With this in mind, we return to Langton's concept of the information ecology (1992). If we accept the premise that living systems have this ecology of information as their basis, then can we also argue that information-processing machines may also lead to the development of living systems? Following Newell and Simon's lead, Silverman and Bullock (2004) offer a PSS Hypothesis for 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. (Silverman and Bullock 2004, p. 5)

Thus, assuming that the computer in question was appropriately programmed to take advantage of this property, then an Alife simulation could be regarded as a true living system, rather than an Artificial<sup>2</sup> imitation of life. As in Newell and Simon, the computer becomes a system that is a sort of blank slate, needing only the appropriate programming to become a repository for digital life.

This PSS Hypothesis handily removes the gap between the living systems which provide the inspiration for Alife and the systems produced by an Alife simulation. Given that computers inherently provide an information ecology, an Alife simulation can harness that property to create a living system within that computer. Strong Alife then becomes a means for producing entirely new datasets derived from genuine digital lifeforms rather than simply a method for creating behaviours that are reminiscent of natural life.

### ***3.6.7 The Importance of Backstory for the Modeller***

As discussed in earlier sections about examples of the theoretical backstory in empirical disciplines, the presence of such a backstory allows the experimenter to justify the usefulness of serious alterations to the system under study. In the case of rat neuroscience and TMS research, their back-stories allow them to describe the changes and preparations they make to their subjects as necessary means for data collection.

In the case of Alife, such a claim is difficult to make, as noted previously, given that the simulation is creating data rather than collecting it from a pre-existing source as is the case in our empirical examples. The PSS Hypothesis for Life avoids this difficulty by stating that any given computer hardware substrate forms the necessary

raw material for life; in a sense, our simulation is merely activating this potential to create an information ecology, and then collecting data from the result. Thus, we may say that our programming of the simulation has prepared this digital substrate for empirical data-collection in a manner similar to that of empirical studies.

Of course, such a backstory is significant in scope, and could invite further criticism. However, such a statement is difficult to refute, given the unsophisticated nature of our current definitions for biological life, and a lack of agreement on whether life may exist in other substrates. Either way, the importance for the modeller is that such a simulation takes on a different character. The focus of such a theoretical justification is on implementing a model which produces some empirical data collection that may bear on our overall understanding of life, not on simply tweaking an interesting computational system to probe the results. The PSS Hypothesis for Life gives us some basis on which to state that this view of Alife models has some theoretical validity.

### ***3.6.8 Where to Go from Here***

Now that we have produced a theoretical backstory which underwrites Alife research as a means for generating new empirical data points, what is missing? The PSS Hypothesis for Life allows us to pursue simulations of this type as a means for understanding life as a new form of biology, investigating the properties and behaviours of entirely novel organisms. One can imagine how such endeavours could provide interesting insight for those interested in the broader questions of what makes something alive.

However, none of this allows us to provide any direct relevance in our results in Alife to the biological sciences. The biologist who views our bird migration model, justified as an information ecology under the PSS Hypothesis for Life, as interesting, but irrelevant to the concerns of the empirical ornithologist. Indeed, if our simulation is producing merely a bird-like manifestation of digital life, then we have no basis on which to state that what we see in our results can tell us anything about real birds, completely removed from the virtual information ecology we have implemented.

This issue becomes the focus of the next chapter, the final chapter of Part I. How might we apply Alife modelling techniques and agent-based models to the broader canvas of biological science? An in-depth discussion of methodological and theoretical issues related to modelling in population biology will provide us with the context necessary to begin to answer this question. The concerns of the strong Alife modeller as depicted in this chapter differ in several important respects to those of a modeller with a weak Alife orientation, and those concerns focus largely on creating a relationship to external empirical data rather than creating their own empirical data as the PSS Hypothesis for Life describes.

### 3.7 Summary and Conclusions

ALife by its very nature is a field which depends upon the use and acceptance of artificially-generated data to investigate aspects of living systems. While some members of this research community have posited that artificial systems can be alive in a manner identical to biological life, there are numerous philosophical and methodological concerns inherent in such a viewpoint.

The comparison with artificial intelligence laid out in this chapter illustrates some of the particular methodological difficulties apparent in ALife. ALife seeks to replicate properties of life which are heavily dependent on the biological substrate, in contrast with AI, which seeks to emulate higher-level properties of living organisms which seem more easily replicable outside of the biological realm.

AI does not entirely escape the problem of artificiality in its data and methods however, nor for that matter does conventional empirical science. Some disciplines are able to make use of a great deal of artificial data by tying it to a theoretical ‘backstory’ of a sort. ALife up to now has lacked this backstory, while AI has Newell and Simon’s PSS Hypothesis (Newell and Simon 1976) as one example of such a theoretical endeavour.

Silverman and Bullock (2004) used Newell and Simon’s account as an inspiration for a PSS Hypothesis for life, a framework which could be used to underwrite strong ALife as a form of empirical endeavour. By accepting that life is a form of information ecology which does not depend exclusively on a biological substrate, a researcher signing up to this backstory may argue for the use of strong ALife data as a form of empirical data. Initially such an idea is appealing; after all, the strong ALife proponent seeks to create forms of ‘life-as-it-could-be’ in a digital space, and this variant of the PSS Hypothesis describes such instances of digital life as a natural consequence of life itself being an information ecology.

However, this framework does not account for the more pragmatic methodological concerns which affect modelling endeavours of this type, and in fact modelling in general. Constructing a computational model requires a great deal more than a useful theoretical justification to function appropriately and provide useful data. As one example, accepting ALife simulations as a form of empirical enquiry does not simplify the task of relating that data to similar data derived from entirely natural systems. As exciting as the prospect of digital life may be, creating self-contained digital ecosystems which are unrelatable to natural ecosystems seems of limited utility for the biologist. Such issues, examined in the context of mathematical models in population biology in the following chapter, will provide greater insight into these crucial elements of our growing theoretical framework for simulation research. This framework can then be expanded to bear on our upcoming discussion of simulation for the social sciences in the second overall section of this text.

## References

- Bedau, M. A. (1998). Philosophical content and method of artificial life. In T. W. Bynum, & J. H. Moor (Eds.), *The digital Phoenix: How computers are changing philosophy* (pp. 135–152). Oxford/Malden, MA: Blackwell Publishers.
- Bennett, B., Magee, D. R., Cohn, A. G., & Hogg, D. C. (2004). Using spatio-temporal continuity constraints to enhance visual tracking of moving objects. *European Conference on Artificial Intelligence*.
- 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 J. Myopoulos, & R. Reiter (Eds.), *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence* (pp. 569–595). San Mateo, CA: Morgan Kaufmann.
- Brooks, R. A., Breazeal, C., Marjanovic, M., Scassellati, B., & Williamson, M. M. (1999). The Cog project: Building a humanoid robot. In Christopher L. Nehaniv (Ed.), *Computation for metaphors, analogy, and agents*. Berlin: Springer.
- Chalmers, A. F. (1999). *What is this thing called science?* Buckingham: Open University Press.
- Churchland, P. M., & Churchland, P. S. (1990). Could a machine think? *Scientific American, 262*, 32–37.
- Dee, H., & Hogg, D. (2006). Navigational strategies and surveillance. In *Sixth IEEE International Workshop on Visual Surveillance*, Graz.
- Di Paolo, E. A., Noble, J., & Bullock, S. (2000). Simulation models as opaque thought experiments. In M. A. Bedau, et al. (Eds.), *Artificial life VII* (pp. 497–506). Cambridge, MA: MIT Press.
- 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*. New York: W.W. Norton.
- Harnad, S. (2005). What is consciousness? *New York Review, 52*(11).
- Keeley, B. L. (1997). Evaluating artificial life and artificial organisms. In C. G. Langton, & K. Shimohara (Eds.), *Artificial life V* (pp. 264–271). Cambridge: MIT Press.
- Kujala, T., Alho, K., & Naatanen, R. (2000). Cross-modal reorganization of human cortical functions. *Trends in Neuroscience, 23*, 115–120.
- Langton, C. G. (1992). Life at the edge of chaos. In C. G. Langton, C. Taylor, J. D. Farmer, & S. Rasmussen (Eds.), *Artificial life II* (pp. 41–91). Reading: Addison-Wesley.
- Langton, C. G., et al. (1989). *Artificial life*. Redwood City: Addison-Wesley.
- Miller, G. F. (1995). Artificial life as theoretical biology: How to do real science with computer simulation. *Cognitive Science Research Paper 378*, University of Sussex, Brighton.
- 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.
- Ray, T. S. (1996). An approach to the synthesis of life. In M. A. Boden (Ed.), *The philosophy of artificial life* (pp. 111–145). Oxford: Oxford University Press.
- Searle, J. R. (1980). Minds, brains and programs. *Behavioral and Brain Sciences, 3*(3), 417–457.
- Searle, J. R. (1982). The chinese room revisited. *Behavioral and Brain Sciences, 5*, 345–348.

- Silverman, E., & Bullock, S. (2004). Empiricism and artificial life. In J. Pollack, M. Bedau, P. Husbands, T. Ikegami, & R. A. Watson (Eds.), *Artificial Life IX: Proceedings of the Ninth International Conference on the Simulation and Synthesis of Living Systems*. MIT Press/Bradford Books.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 49, 433–460.
- Winograd, T. (1972). Understanding natural language. *Cognitive Psychology*, 3(1), 1–191.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

