

Structuring Intelligence: The Role of Hierarchy, Modularity and Learning in Generating Intelligent Behaviour

Joanna J. Bryson
Artificial models of natural Intelligence
University of Bath, United Kingdom

March 29, 2010

1 Introduction

Scientific social trends ranging from dynamical systems theories to post-modernism have called into question whether the apparent hierarchical structure of naturally-occurring intelligent behaviour actually derives from structured intelligence. These questions and perspectives overlook substantial evidence from both neuroscience and biology more broadly that behaviour really is organised utilising both modularity and hierarchy. In mammal brains, modular structure starts from cellular composition, and continues conspicuously through the existence of discrete regions with differing processing capacities (Badre, 2008). We can discriminate a brain region by its consistent and regular pattern of nerve-cell type and inter-cell connectivity, while these same features vary between regions. Hierarchy derives from the interaction between these modules. This is particularly apparent when we can produce complex movements such as the production of particular words (Mateer, Rappoport II, & Polly, 1990) or grasping and transfer gestures (Graziano, Taylor, & Moore, 2002) by directly stimulating single cells.

Computer science tells us that a wide variety of possible computational architectures are able to produce the same computational outcome (Turing, 1936). Why then would evolution select something so elaborate as a highly-regionally-differentiated vertebrate brain? As scientists we believe that the most parsimonious explanation is the most probable. Similarly, the most simple structure sufficient is the most likely to be discovered and maintained by evolution. However, recognising parsimony is not always easy (Dawkins, 1997; Myung, Forster, & Browne, 2000). Many factors need to be taken into consideration, such as path dependency (the impact of historical accident) and metabolic efficiency.

Theoretical computer science can help us understand why evolution reliably finds structured solutions. Efficiency is determined by task, in learning and information just as in physics (Wolpert, 1996). A modular brain addresses the variety of tasks intelligence requires with an efficiency that no single representation could provide. And animal-like intelligence does require varied capabilities. The most obvious example of this is the variety of sensory inputs we utilise. Light and sound each provide us with information the other cannot — the same is true of somatic touch and internal proprioception. The sensory organs that first detect these forms of information can already be thought of as specialised modules of our overall intelligence, but so too could the various cortical regions that process this information further downstream. Here the specialist regions may not be specialised just to sensory modality, but rather to information structure. For example, the “visual cortex” in fact processes spatial information more generally — it is also invoked in learning Braille (Kauffman, Theoret, & Pascual-Leone, 2002).

Following from the above brief introduction to the evidence of and explanations for structure in natural intelligence, this chapter will next describe the history of ideas of structured control, attempting to explain their periodic losses and resurgence. I then address a few of the criticisms that underly current skepticism concerning structured control. Finally, I discuss the origins of intelligent behaviour in the large, showing that modularity and hierarchy are pervasive in explaining the varieties of intelligent behaviour found in nature.

2 The History of Theories of Structured Control

We often tend to think of scientific understanding as a monotonically increasing library of information, but this is not the case. While we may accumulate data more or less like a collection, real understanding requires the construction of theories. This construction is a social process. Scientists are trained to know that science is never about certainty. Science is rather a dynamic process concerned with *approaching* certainty through evidence and reason. Thus at any particular time, leading experts on a topic will know a variety of hypotheses, the evidence for and against each of them, and experts associated with each of these matters. *Scientific understanding* then is really an aggregate term, reflecting a variety of dominant theories and individual hypotheses held at a particular time.

In this section I review the last century of history concerning the scientific evidence for and against structure playing a role in intelligent control. Given, as the previous section summarised, the prevalence of evidence for structure in intelligence, understanding why there is any doubt in the matter requires understanding the importance of theory on understanding, and the importance of personality, philosophy, politics and history on theory.

2.1 Structure

One of the defining features of intelligent behaviour is the ordering of expressed individual actions into coherent, apparently rational patterns. From approximately 1951 until the mid 1980s, the dominant theories in both psychology and artificial intelligence for explaining intelligent behaviour held that hierarchical and sequential structures internal to the agent or animal underlie this ordered expression (e.g. Lashley, 1951; Tinbergen, 1951; Piaget, 1954; Hull, 1943; Dawkins, 1976; McGonigle & Chalmers, 1996). However, the last two decades have seen an increase of support for a more dynamic theory of intelligence (e.g. Port & van Gelder, 1995; Botvinick, 2007). This new theory holds that intelligence, like the brain itself, is actually composed of enormous numbers of small processes operating in parallel. Several researchers in this new paradigm have claimed that behaviour controlled by hierarchy is necessarily rigid, brittle, and incapable of reacting quickly and opportunistically to changes in the environment (Maes, 1991; Goldfield, 1995; Hendriks-Jansen, 1996; Seth, 2007). They suggest that the apparent hierarchical organisation of behaviour is not the result of internal structured control, but it is rather only an inadequate model imposed on a far more complex dynamic process.

Hendriks-Jansen traces the hierarchical theory of behaviour organisation to the ethologist McDougall (1923), who presented a theory of the hierarchy of instincts. Ethological theory during this period, however, was dominated by Lorenz, who “denied the existence of superimposed mechanisms controlling the elements of groups” instead believing that “the occurrence of a particular activity was only dependent on the external stimulation and on the threshold for release of that activity.” (Baerends 1976 p. 726 cited in Hendriks-Jansen 1996 pp. 233–234). This theory had obvious correlates with the then-popular black-box theories of Skinner (1935) and the behaviourists, though these were actually agnostic as to the underlying organisation of behaviour. The behaviourists’ agnosticism was itself a scientific-political position set in opposition to Freud.

Lashley (1951) revitalised the hierarchical theory of behaviour with both data and reason. He demonstrated that hierarchy is the only explanation for the speed of some action sequences, such as those involved in human speech or the motions of the fingers on a musical instrument. Neural processes are simply too slow to allow elements of such sequences to be independently triggered in response to one another. Lashley therefore proposes that all the elements of such a sequence must be simultaneously activated by a separate process — the definition of hierarchical organisation. The exact neurological explanations of sequence learning are still an active area of research. However, Lashley's argument is well established. Items of a sequence are initially activated hierarchically, as a set, and then released by a second mechanism (Salthouse, 1986; Henson, Norris, Page, & Baddeley, 1996; Davelaar, 2007).

From roughly the time of Lashley's analysis, hierarchical models have dominated attempts to model intelligence. Particularly notable are the models of Tinbergen (1951) and Hull (1943) in ethology, Chomsky (1957) in linguistics, and Newell and Simon (1972) in artificial intelligence and human problem solving. Mainstream psychology has been less concerned with creating specific models of behaviour control, but generally assumes hierarchical organisation as either an implicit or explicit consequence of goal-directed or cognitive theories of behaviour (Bruner, 1982). Staged theories of development and learning are also hierarchical when they described complex skills being composed of simpler, previously-developed ones (Piaget, 1954; Greenfield, 1991; Karmiloff-Smith, 1992).

In a tribute to Tinbergen (his PhD supervisor), Dawkins (1976) extended Lashley's argument for hierarchical control. Dawkins used computational parsimony to argue *for* hierarchical theories. Dawkins argues that it is more likely that a complex action sequence useful in multiple situations should be evolved or learned a single time, and that it is also more efficient to store a single instance of such a skill.

Dawkins' arguments and proposals anticipate many of the techniques developed later for robust control in artificial intelligence. In particular, Brooks (1986, 1991b) advocates the decomposition of intelligence into task-specific modules. These modules, termed *behaviours* are in turn organised into fairly egalitarian networks of competence, termed *levels*. As the name implies, the levels are themselves organised in a stack, with the highest level subsuming the goals of the lower levels, though these are assumed to be operating in parallel. For each level, the levels below it provide a reliable behavioural context for its operation. Ironically, although Brooks' system, the Subsumption Architecture, exploits both modular organisation and a simple, linear, goal-based hierarchy of module coordination, his approach has been taken by journalists and philosophers as evidence for homogeneous dynamical models intelligence (Kelly, 1995; Hendriks-Jansen, 1996). Shanahan (2005) points out a similar inconsistency in some people's perception of the Global Workspace Theory of conscious action selection.

More recently, many authors treat the hierarchical and modular structure of intelligence as obvious and accepted facts (Byrne, 1999; Prescott, Bryson, & Seth, 2007). Nevertheless, the controversy persists (Barsalou, Breazeal, & Smith, 2007; Botvinick, 2007; Seth, 2007).

2.2 Emergence

The competing theory, that responsive animal intelligence cannot possibly be governed by hierarchical control, has emerged from some of the practitioners of the dynamic hypothesis of cognition (van Gelder, 1998). These researchers tend to seek a single mathematical system of representation to be sufficient for describing and learning intelligent behaviour. Not all researchers that fit this description are in-principle opposed to hierarchical structure. For example, since the mid 1990s, much machine learning has exploited *hierarchical* hidden Markov models (Fine, Singer, & Tishby, 1998). Although these are provably equivalent to single-layer Markov models, learning systems are simply easier to engineer when hierarchy is added. The same is true in neural network architectures, though here there are theoretical reasons to have at least two layers (Minsky & Papert,

1969). Nevertheless, most researchers choose at least three, for purely practical engineering reasons (Hansen & Salamon, 1990). Such approaches though are still relatively homogeneous, overlooking the example of nature favouring modularity as well as hierarchy.

The theory of dynamic action expression suggests that complex dynamic or chaotic systems operate within the brain producing the next behaviour not by selecting an element of a plan, but rather as an emergent consequence of many parallel processes (e.g. Maes, 1991; Brooks, 1991a; Goldfield, 1995; Kelso, 1995; Hendriks-Jansen, 1996; van Gelder, 1998; Barsalou et al., 2007; Seth, 2007). Evidence supporting the older hypothesis of structured hierarchical behaviour is seen to have been biased by the hierarchical and sequential nature of human explicit thought and language. In particular, because much theoretical work in psychology is conducted using computer models, theories may be biased towards the workings and languages of the serial processors of the machines available to most psychologists (Brooks, 1991a).

A fundamental appeal of the dynamic hypothesis is that it is necessarily correct, at least to some level. Assuming a materialist stance, intelligence is known to be based in the parallel operation of the body's neural and endocrine systems. It is nearly as well accepted that human and animal behaviour can be described as hierarchically ordered (Dawkins, 1976; Greenfield, 1991; Byrne & Russon, 1998; Prescott et al., 2007). Though note even here that in response to the Byrne and Russon (1998) target article on imitation of the hierarchical structure of behaviour, nearly a fifth of the commentaries chosen to appear with the article questioned the existence of hierarchical control (Vereijken & Whiting, 1998; Mac Aogáin, 1998; Jorion, 1998; Gardner & Heyes, 1998, of 21 commentaries).

The question is, given that there is statistically discriminable sequential and hierarchical ordering of behaviour, how and when are these behaviours so organised? Some argue that the order is *emergent*, which is to say not encoded explicitly but rather a consequence of the interaction of a system of simpler processes. Viewed from an atomic or even cellular level, this is of course necessarily true of all behaviour — in fact, of all matter. Further, few would argue that animals have complete structured behaviour plans that are expressed regardless of their environment. However, the structure of actions we see emerges not from an amorphous homogeneous being, but from the modular and hierarchical neural control system found in vertebrates. It is natural and parsimonious for us to conclude that this structure at least partially determines the patterns of action it generates.

3 The Confusion of Theories of Structured Control

Given the strong evidence from neuroscience for structured intelligent control, why does controversy continue in this area? There are several possible reasons. First science is a political process carried out by humans with deep personal concerns and values. In particular, the twentieth century witnessed a large swing in human thought towards the importance of egalitarianism. Such political influence has had documented effect on evolutionary theory (Sterelny, 2007) — it may have influenced ethology as well. Second, as the mathematics of chaos and complexity became known, it was correct scientific procedure as well as natural that its practitioners should try to account for as much data as possible with this new approach. But finally, there have been a few specific influential theories that ultimately proved to be simply wrong. These wrong theories have turned some scientists away from concepts associated with the theories. In this section I address a few of these misconceptions.

3.1 Modularity Does Not Require Language, Strict Encapsulation or Innateness

One reason scientists such as Elman et al. (1996) or Barsalou et al. (2007) reject modularity as an explanation of human intelligence is because they are really rejecting one particular characterisation

of it. In psychology, the best known description of modularity is due to Fodor (1983). One of the primary criteria Fodor provides for recognising a module is that a module must be innate, not learned. The lifelong interplay between genes and environment makes many developmental psychologists and biologists uncomfortable with the entire category of innateness (e.g. Thelen & Smith, 1994; Elman et al., 1996; Donnai & Karmiloff-Smith, 2000; Griffiths, 2001). Of course, innateness has no bearing on the functional or computational characteristics of modularity. Many proponents of psychological modularity believe that modules develop over the ontogeny of an organism (e.g. Bates, 1999; Carruthers, 2005; Sperber & Hirschfeld, 2006). Here the idea is that regions or subnetworks of the brain specialise to particular tasks in a response to experience, in an ontogenetic recapitulation of the evolution of the neural specialisation we see in the brain's various regions.

Similarly, Barsalou et al. (2007) reject modular theories because modules must be fully encapsulated — that is, they cannot affect each other at the level of data or representation. This characterisation is also due to Fodor (1983), though in AI it was particularly popularised by Brooks (1991b). Not every modular system shares this characteristic either, the brain being an obvious exception. But in AI too, memory can be usefully stored in modules, possibly with special-purpose representations, yet still made accessible to other modules via interfaces (Bryson, 2000a, 2001; Thórisson, Pennock, List, & DiPirro, 2004; Metta, Fitzpatrick, & Natale, 2006).

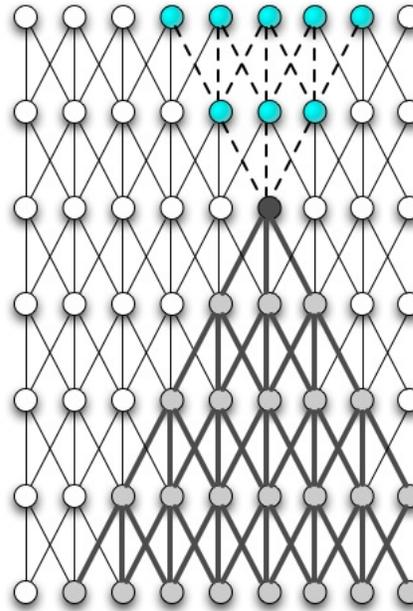
Another confusion that Fodor (1983) introduced is the claim that modules only support perception and action, and must be combined into a true mind via a layer composed of general intelligence that translates between them. This somewhat bizarre claim has not only been accepted but modified and extended to imply that modules can only be integrated with language, and that this explains human exceptionality (Carruthers, 2003; Spelke, 2003; Chomsky, 2000). While there is no question that language substantially affects human intelligence and culture, as has already been explained modular systems exist throughout natural and artificial intelligence, in systems well-integrated without language. Human exceptionalism in accumulating culture probably substantially predates language, and in fact must account for language itself (Buckley & Steele, 2002; Bryson, 2009).

3.2 Hierarchical Organisation Does Not Imply Slow, Autocratic or Fragile Control

In the introduction I mentioned that the presence of hierarchical control in the brain was demonstrated in neuroscience by the production of complex movements by directly stimulating single cells. Examples include speaking particular words in humans (Mateer et al., 1990) and making grasp and transfer gestures appropriate to scramble feeding in macaques (Graziano et al., 2002). Considering such cells to be near an apex of a pyramid of top-down control is the classic error of reasoning that has motivated much anti-hierarchical rhetoric. Rather, such cells should be thought of as a juncture, an intersection between two overlapping hierarchies (see Figure 1). The first hierarchy generates behaviour when the cell is highly stimulated. The second hierarchy is the combination of motivation and perceived opportunity that would ordinarily bring activation and other information to the cell when neuroscientists were not the ones stimulating it. In terms of flow of information through the brain, the two hierarchies can be seen as inverted from one another (see figure). Neighbouring cells in the same layer of the hierarchy receive nearly the same set of inputs and distribute nearly the same sort of outputs. These neighbouring cells will generally be interconnected to each other through patterns of mutual inhibition and excitation, simplifying the coordination problem between cells at the same level (Riesenhuber & Poggio, 1999).

The firing of each such cell is sufficient to generate behaviour involving the the further processing of millions of cells — not only is the action executed by thousands of muscle cells signalled by thousands of afferent nerves, but also the basic signal is further modulated and controlled by brain structures such as the cerebellum and basic feedback control loops in the spine. But what causes these cells to fire in the first place is at least as complex — a system of not only perception and sen-

Figure 1: A gross abstraction of neural control. Hierarchical organisation can occur within columns, where peer nodes on the level of hierarchy each specialise as an apex for processing a certain set of sensory information, and for generating a certain form of behavioural control. Moderation between nodes at the same level of hierarchy happens both through mutual excitation and inhibition, and through executive control (c.f. Redgrave, Prescott, & Gurney, 1999).



fusion, but also attention and goal arbitration. The normal stimulus for such an action requires recognition of external stimuli, itself a hierarchical process fusing information from the many cells that compose sensory apparatus with learned categories. Any normal environment contains many such recognisable stimuli, and which one an animal attends to must itself be governed in a way appropriate to the animal's current needs (Redgrave, Prescott, & Gurney, 1999).

The fact that the brain is essentially concurrent also allows an explanation for how hierarchically-structured behaviour can avoid the fragility normally associated with single-point control, and also avoid latencies associated with having to revisit every step of a hierarchy (a criticism of hierarchy due to Maes, 1991). Forebrain gating systems may focus attention directly on computation at one subtask, even though a hierarchically-organised path brought attention to that task in the first place (Rensink, 2000; Bryson & McGonigle, 1998). The sort of distributed system described above and shown in Figure 1 is also robust to damage, as such networks adapt dynamically to account for data — they are perpetually learning in order to make more reliable predictions. Thus if damage removes part of the network, neighbouring cells adapt to fill in the gap in control.

My own work has shown in artificial intelligence that hierarchy working in interaction with concurrent modules can control mobile robots navigating ordinary office space (Bryson & McGonigle, 1998) and play complex real-time computer games like Capture the Flag against human opponents (Partington & Bryson, 2005). Further, I have shown that such hierarchical structure with switching attention can out-perform a 'fully aware' system at controlling an agent with multiple goals and threats in a fairly rich environment (Bryson, 2000b). In this last case I argue that my system worked better purely because it was easier to design and to adjust. Because the hierarchy reduced the complexity of the search space the designer has to parse, the program is sufficiently easier to optimise that this compensates for any loss of information through neglect of available options. The same combinatorics confront evolution and development in biology as confront human designers. Thus hierarchical structures provides advantages in all three cases.

4 Modularity in Control Extends Beyond Intelligence

So far I have described modularity as a decomposition of intelligent control for the purpose of reliability and / or efficiency. A sufficient though not necessary characteristic of modularity is for each module to have a distinct system of representation for storing acquired information. Once we begin to think about control as being decomposed modularly, we can also begin extending our understanding of intelligence into an understanding of the biological and evolutionary origins of cognition. This in turn can help us understand species-level differences in cognition and related traits, such as life history.

Modularity is already argued outside of psychology and neuroscience in evolutionary biology. Evolutionary developmental biology argues that physiological modularity, encoded by the genome, underlies and accelerates contemporary tetrapod evolutionary processes (Kirschner, Gerhart, & Norton, 2006; Müller, 2008). Single mutations do not tend to make miniscule alterations to an animal, but rather to add, subtract or alter complexes of interrelated traits. For example, the shape of the beaks of Darwin's finches on the Galapagos is controlled by only two genes, and the same genes account for general head shape, skull thickness etc. so that a bird will always be able to support its own beak (Campàs, Mallarino, Herrel, Abzhanov, & Brenner, 2010). My own work assigns exactly this role to modularity in the development of AI systems. Plasticity should only be made available where intelligence cannot practically be provided in advance by the developer, for example if a robot will be required to learn its own navigation routes in unknown buildings. Where such learning must take place, it should be performed in purpose-built modules that support the learning to the maximum extent possible (Bryson, 2001, 2003).

Can reasoning about modularity give insight into species-level differences in cognition? What needs to be explained? Cognition is often portrayed as being a universally useful ability. If this is true, we need to explain the diversity of cognitive capacities. We may be tempted to take an implicitly Lamarckian view, thinking that somehow humans are just particularly lucky or fittest, but in fact the discrepancies in fitness between extant species are not great (Kimura, 1985). Rather, extant species represent a position in a space of possible evolutionary tradeoffs. The best-known example of such a tradeoff is the quantity of offspring versus the amount of parental investment per offspring. We can think of species as ranged along an axis of variation between investing a great deal in a very small number of offspring (as with juvenile apes) or very little in a large number of offspring (as with fungus spores). Extant variation in cognitive abilities, ranging from apes through protozoa, indicates that there must be some compensating tradeoff or tradeoffs against cognition, creating axes of cognitive variation.

One of the axes of cognitive variation is the speed with which learning takes place. Even within species, the vertebrate brain consists of modules which learn at a variety of rates. Quick-learning systems can rapidly gather data from direct experience, but only because they can recognise and represent that information with sparse representations (Teyler & Discenna, 1986; McClelland, McNaughton, & O'Reilly, 1995). The categories necessary for these sparse representations are formed through slower systems which extract generalisations from large quantities of data, possibly considering repeated presentations (Louie & Wilson, 2001). This is one of the standard theories of the computational function of the hippocampal system — that it quickly gathers episodic memories which can then be used to train cortical systems on semantic knowledge. This episodic data itself relies on cortical / semantic category representation to provide a sufficiently sparse representation for storing large amounts of information in a confined brain region (Teyler & Discenna, 1986; McClelland et al., 1995; Louie & Wilson, 2001; Rogers & McClelland, 2004; Foster & Wilson, 2006).

Not every aspect of individual behaviour is learned. A great deal of behaviour is determined genetically, including (as the behaviourists eventually discovered) which sorts of stimuli can be associated with which types of action (Gallistel, Brown, Carey, Gelman, & Keil, 1991). This is true not only from the perspective of the architecture of the brain, but also the morphological characteristics of the body, which determine the possible range of an animal's actions, its capacities of perception,

and its metabolic needs (Nehmzow, Smithers, & McGonigle, 1993; Pfeifer, 1999; Paul, 2004). Genetic evolution might also be viewed as a process of learning — a means by which species and therefore individuals within species acquire behaviour.

If we are willing to think of learning (or at least behaviour acquisition) in this way — as something that sometimes occurs outside of an individual — then we can postulate a continuum of adaptation processes running from individual development and learning through evolution and speciation. The computational costs and benefits of modularity and hierarchy are the same in all these cases. Thus we find modular architectures throughout nature: interacting but very different representations that generate behaviour. Genes in a genome, nerve cells in a brain, individuals in a society, species in an ecosystem — each of these levels of hierarchy has particular mechanisms for producing behaviour, and of acquiring and representing variation, and of performing optimisation across the possible behaviours. Each layer itself contains modules composed of relatively more homogeneous individual elements expressing relatively similar strategies across a relatively narrow spread of variation. For example, we can see in tetrapod embryos stable selection for two different types of development: a robust core spinal area where each cell is autonomous and controls its own expression, and the elaborated limb regions where hierarchical control of gene expression allows the relatively rapid evolution of novel forms (Winslow, Takimoto-Kimura, & Burke, 2007). Neither representation usurps the other, though one (the spinal version) is older. The hierarchical representation increases evolvability so is adaptive, but also depends on a more robust representation at the core. Within these modules, further sub-modules develop which become limbs, digits, and organs (Müller, 2008).

We can similarly think of species as modules within an ecosystem. If all currently-present life on Earth has evolved from a single replicator, then the fact that the subsets of individuals that engage in recombination of their genotypes is not entirely arbitrary is interesting. Which couples co-parent offspring depends not only on physical proximity, but also on species. Thus a species might be thought of as a modularised search mechanism, which optimises a particular behavioural strategy which is encoded in that species' genotype.

One interesting tradeoff helps bridge two of the levels I described earlier — the individual vs. the society. This is the inclusive-fitness tradeoff. This tradeoff determines how much effort an individual gives to its own reproductive success vs. its relatives' (see e.g. West, Griffin, & Gardner, 2007). For example, one is twice as related to one's own children than to one's nieces or nephews, so inclusive fitness tells us that we should be twice as ready to invest effort in our own children as sibling's (Hamilton, 1964). Of course, breeding can incur significant costs. Normally, one is not considered to be altruistic if one invests effort in one's offspring, but is this simplistic definition really justified? Altruism is not a mystery of evolution, but rather a measure of where along a spectrum of potential inclusive-fitness investments a species happens to lie. Individuals from more altruistic species invest more heavily in the reproductive success of their extended kin, though still presumably the cost is in proportion to their kin's benefit and relatedness.

In humans and other cognitive social species, behaviour is acquired not only from genetics but from culture — from social transmission. We might choose to define *kin* in terms of sharing any inherited traits which control the inclusive-fitness tradeoff, whether biological or cultural. This leap allows us to generalise our explanation of altruism further. Levels of individual altruism may reflect not only species-level strategies, but also context- or personality- based variation in individual strategies. For example, Rutte and Taborsky (2007) have demonstrated that individuals of some species behave more altruistically when they experience or even witness more altruistic behaviour themselves. This presumably facilitates more rapid adaptation to social vs. individual strategies, in response to the current environmental context.

Where levels of altruism are personality based, the distribution of such personalities in a population is still subject to selection. Where they respond to recent individual experience, this is an example of a tradeoff being adjusted 'on the fly' based on individual experience. Of course, even where phenotypic plasticity is exploited this way, the range of possible individual variation and

the mechanisms for assessing appropriate levels of investment are also subject to selection.

To return to our starting point, there must also be a set of fitness tradeoffs that relate to the expression of cognition itself. Here too the tradeoffs concern the intersection of two of my proposed levels for representing behaviour. Cognition is a special case of individual plasticity, and the tradeoffs between individual plasticity and biological evolution are beginning to be well understood (Hinton & Nowlan, 1987; Maynard Smith, 1987; Borenstein, Meilijson, & Ruppin, 2006; Paenke, Sendhoff, & Kawecki, 2007). Individual plasticity can accelerate biological evolution, but biological evolution provides greater reliability than individual learning. Thus in a stable environment, selection generally favours the reduction of individual plasticity, but in periods of rapid environmental change (including where that change is caused by ecological dynamics e.g. in the number or proportion of species), there will be more pressure for individual plasticity. Interestingly, even in stable environments, the presence of individual plasticity can sometimes *decelerate* evolution as the genome becomes very near the optimum. This is because learning at this point becomes sufficiently reliable that there is no selective pressure to eliminate the final amount of variation from the optimum. Thus phenotypic plasticity may be a key factor in maintaining sufficient genetic variation in the population to allow robustness in the case that a previously-stable selective environment suddenly does change.

Tradeoffs concerning cognition occur between species as well as within them (Lefebvre, Reader, & Sol, 2004; Barrickman, Bastian, Isler, & van Schaik, 2008). Species that commit to highly cognitive strategies may compensate for the uncertainty of individual learning by providing epigenetic information transmission such as social transmission of behaviour. Species that rely on social transmission though pay a penalty not only in terms of metabolic cognitive cost, but also in terms of extended life history which decelerates the rate of biological evolution. Culture could in theory allow species to take advantage of massive increases in both the rate of acquisition of new information and the reliability of learning due to concurrency — the ability to exploit the cognitive capacity of many individuals simultaneously (Bryson, 2009). However in practice only humans seem to exploit this capacity fully and then only in very-recent history from an evolutionary perspective (Ambrose, 2001).

The main constraints on using culture as a source of behaviour have to do with the probability of transmitting information into the next generation. This is determined by a number of factors, such as the rate of information transmission between individuals, the amount of time individuals spend in proximity, and the average and maximum lifespans of the individuals (Bryson, Bilovich, & Čače, 2009). Of course, not every agent needs to learn all behaviours in a culture. There must only be a sufficient number of individuals in the population carrying a culture that a particular behaviour is reliably carried forward from generation to generation by at least some individuals. The probability of social learning can be even lower if the probability of individual discovery of the behaviour is sufficiently high to compensate.

In this section then I have argued that modularity can be seen as a useful organisational principle for developing adaptive behaviour across a wide range of representations. These representations — genes, nerve cells, individuals, societies, species — each support adaptation, and have a variety of mechanisms for operating within modules, between modules, and between the layers of hierarchy determined by the representational substrate. Neural or psychological modularity such as the majority of this chapter has discussed is then just a special case of generally useful organisational strategies.

5 Conclusion

My argument in this chapter has been that hierarchical structures, including modularity, are pervasive in natural intelligence — though so too are concurrent solutions. The reason for this is simple combinatorial complexity. Control and design are more easily addressed when broken into

units, both as parallel modules and as layers of control. AI researchers often neglect the hierarchical nature of their own solutions, as they build homogeneous networks for control, yet keep these networks small and simple while having the items that they control be well-designed powerful operators. Similarly some natural scientists argue vociferously against modular models of mind while never doubting that the brain, the limbs and the sense organs operate in different ways and make different contributions into intelligent behaviour. In this chapter I have illustrated how modularity and hierarchy pervade intelligence, from genetic evolution through the neural structure of vertebrate brains, the accumulation of social learning by a society, and the differentiation of species' behaviour in an ecosystem.

References

- Ambrose, S. H. (2001). Paleolithic technology and human evolution. *Science*, 291, 1748–1753.
- Badre, D. (2008). Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends in Cognitive Sciences*, 12(5), 193–200.
- Baerends, G. P. (1976). The functional organisation of behaviour. *Animal Behaviour*, 24, 726–738.
- Barrickman, N. L., Bastian, M. L., Isler, K. & van Schaik, C. P. (2008). Life history costs and benefits of encephalization: A comparative test using data from long-term studies of primates in the wild. *Journal of Human Evolution*, 54(5), 568–590.
- Barsalou, L. W., Breazeal, C. & Smith, L. B. (2007). Cognition as coordinated non-cognition. *Cognitive Processing*, 8(2), 79–91.
- Bates, E. (1999). Plasticity, localization and language development. In S. Broman & J. M. Fletcher (Eds.), *The changing nervous system: neurobehavioral consequences of early brain disorders* (pp. 214–253). Oxford University Press.
- Borenstein, E., Meilijson, I. & Ruppin, E. (2006). The effect of phenotypic plasticity on evolution in multi-peaked fitness landscapes. *Journal of Evolutionary Biology*, 19(5), 1555–1570.
- Botvinick, M. M. (2007). Multilevel structure in behaviour and in the brain: A model of Fuster's hierarchy. *Philosophical Transactions of the Royal Society, B—Biology*, 362(1485), 1615–1626.
- Brooks, R. A. (1986, April). A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, RA-2, 14–23.
- Brooks, R. A. (1991, August). Intelligence without reason. In *Proceedings of the 1991 international joint conference on artificial intelligence* (pp. 569–595). Sydney.
- Brooks, R. A. (1991b). Intelligence without representation. *Artificial Intelligence*, 47, 139–159.
- Bruner, J. S. (1982). The organisation of action and the nature of adult-infant transaction. In M. von Cranach & R. Harré (Eds.), *The analysis of action* (pp. 313–328). Cambridge University Press.
- Bryson, J. J. (2000a). Cross-paradigm analysis of autonomous agent architecture. *Journal of Experimental and Theoretical Artificial Intelligence*, 12(2), 165–190.
- Bryson, J. J. (2000b). Hierarchy and sequence vs. full parallelism in reactive action selection architectures. In *From animals to animats 6 (sab00)* (pp. 147–156). Cambridge, MA: MIT Press.
- Bryson, J. J. (2001). *Intelligence by design: Principles of modularity and coordination for engineering complex adaptive agents*. (PhD thesis, MIT, Department of EECS). AI Technical Report 2001-003.
- Bryson, J. J. (2003). The Behavior-Oriented Design of modular agent intelligence. In R. Kowalszyk, J. P. Müller, H. Tianfield & R. Unland (Eds.), *Agent technologies, infrastructures, tools, and applications for e-Services* (pp. 61–76). Berlin: Springer.
- Bryson, J. J. (2009). Representations underlying social learning and cultural evolution. *Interaction Studies*, 10(1), 77–100.
- Bryson, J. J., Bilovich, A. & Čače, I. (2009). *Factors limiting the biological evolution of cultural accumulation*. submitted.

- Bryson, J. J. & McGonigle, B. (1998). Agent architecture as object oriented design. In M. P. Singh, A. S. Rao & M. J. Wooldridge (Eds.), *The fourth international workshop on agent theories, architectures, and languages (atal97)* (pp. 15–30). Providence, RI: Springer.
- Buckley, C. & Steele, J. (2002). Evolutionary ecology of spoken language: Co-evolutionary hypotheses are testable. *World Archaeology*, 34(1), 26–46.
- Byrne, R. W. (1999). Imitation without intentionality. using string parsing to copy the organization of behaviour. *Animal Cognition*, 2(2), 63–72.
- Byrne, R. W. & Russon, A. E. (1998). Learning by imitation: A hierarchical approach. *Brain and Behavioral Sciences*, 21(5), 667–721.
- Campàs, O., Mallarino, R., Herrel, A., Abzhanov, A. & Brenner, M. P. (2010). Scaling and shear transformations capture beak shape variation in Darwin's finches. *Proceedings of the National Academy of Sciences*, 107(8), 3356–3360.
- Carruthers, P. (2003). The cognitive functions of language. *Brain and Behavioral Sciences*, 25(6), 657–674.
- Carruthers, P. (2005). The case for massively modular models of mind. In R. Stainton (Ed.), *Contemporary debates in cognitive science* (pp. 205–225). Blackwell Publishing.
- Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.
- Chomsky, N. (2000). Linguistics and brain science. In A. Marantz, Y. Miyashita & W. O'Neil (Eds.), *Image, language, brain: Papers from the first mind articulation project symposium* (pp. 13–28). Cambridge, MA: MIT Press.
- Davelaar, E. J. (2007). Sequential retrieval and inhibition of parallel (re)activated representations: A neurocomputational comparison of competitive queuing and resampling models. *Adaptive Behavior*, 15(1), 51–71.
- Dawkins, R. (1976). Hierarchical organisation: a candidate principle for ethology. In P. P. G. Bateson & R. A. Hinde (Eds.), *Growing points in ethology* (pp. 7–54). Cambridge: Cambridge University Press.
- Dawkins, R. (1997). *Climbing mount improbable*. New York: W. W. Norton & Co.
- Donnai, D. & Karmiloff-Smith, A. (2000). Williams syndrome: from genotype through to the cognitive phenotype. *American Journal of Medical Genetics*, 97(2), 164–171.
- Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D. & Plunkett, K. (1996). *Rethinking innateness. a connectionist perspective on development*. Cambridge, MA: MIT Press.
- Fine, S., Singer, Y. & Tishby, N. (1998). The hierarchical hidden Markov model: Analysis and applications. *Machine learning*, 32(1), 41–62.
- Fodor, J. A. (1983). *The modularity of mind*. Bradford Books. Cambridge, MA: MIT Press.
- Foster, D. J. & Wilson, M. A. (2006). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature*, 440(7084), 680–683.
- Gallistel, C., Brown, A. L., Carey, S., Gelman, R. & Keil, F. C. (1991). Lessons from animal learning for the study of cognitive development. In S. Carey & R. Gelman (Eds.), *The epigenesis of mind* (pp. 3–36). Hillsdale, NJ: Lawrence Erlbaum.
- Gardner, M. & Heyes, C. (1998). Splitting lumping and priming. *Brain and Behavioral Sciences*, 21(5), 695.
- Goldfield, E. C. (1995). *Emergent forms: origins and early development of human action and perception*. Oxford University Press.
- Graziano, M. S. A., Taylor, C. S. R. & Moore, T. (2002). Complex movements evoked by microstimulation of precentral cortex. *Neuron*, 34, 841–851.
- Greenfield, P. M. (1991). Language, tools and brain: the ontogeny and phylogeny of hierarchically organized sequential behavior. *Brain and Behavioral Sciences*, 14, 531–595.
- Griffiths, P. E. (2001). Genetic information: A metaphor in search of a theory. *Philosophy of Science*, 68(3), 394–412.

- Hamilton, W. D. (1964). The genetical evolution of social behaviour. *Journal of Theoretical Biology*, 7, 1–52.
- Hansen, L. K. & Salamon, P. (1990). Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(10), 993–1001.
- Hendriks-Jansen, H. (1996). *Catching ourselves in the act: situated activity, interactive emergence, evolution, and human thought*. Cambridge, MA: MIT Press.
- Henson, R. N. A., Norris, D. G., Page, M. P. A. & Baddeley, A. D. (1996). Unchained memory: Error patterns rule out chaining model of immediate serial recall. *Quarterly Journal of Experimental Psychology*.
- Hinton, G. E. & Nowlan, S. J. (1987). How learning can guide evolution. *Complex Systems*, 1, 495–502.
- Hull, C. (1943). *Principles of behaviour: an introduction to behaviour theory*. New York: D. Appleton-Century Company.
- Jorion, P. J. M. (1998). A methodological behaviourist model for imitation. *Brain and Behavioral Sciences*, 21(5), 695.
- Karmiloff-Smith, A. (1992). *Beyond modularity: a developmental perspective on cognitive change*. Cambridge, MA: MIT Press.
- Kauffman, T., Theoret, H. & Pascual-Leone, A. (2002). Braille character discrimination in blind-folded human subjects. *Neuroreport*, 13(5), 571–574.
- Kelly, K. (1995). *Out of control: The new biology of machines, social systems, and the economic world*. Boston, MA, USA: Addison-Wesley Longman.
- Kelso, J. S. (1995). *Dynamic patterns: the self-organization of brain and behavior*. MIT Press.
- Kimura, M. (1985). *The neutral theory of molecular evolution*. Cambridge University Press.
- Kirschner, M. W., Gerhart, J. C. & Norton, J. (2006). *The plausibility of life*. New Haven, CT: Yale University Press.
- Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior*. New York: John Wiley & Sons.
- Lefebvre, L., Reader, S. M. & Sol, D. (2004). Brains, innovations and evolution in birds and primates. *Brain, Behavior and Evolution*, 63(4), 233–246.
- Louie, K. & Wilson, M. A. (2001). Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. *Neuron*, 29(1), 145–156.
- Mac Aogáin, E. (1998). Imitation without attitudes. *Brain and Behavioral Sciences*, 21(5), 696–697.
- Maes, P. (1991). A bottom-up mechanism for behavior selection in an artificial creature. In J.-A. Meyer & S. Wilson (Eds.), *From animals to animats* (pp. 478–485). Cambridge, MA: MIT Press.
- Mateer, C. A., Rapport II, R. L. & Polly, D. D. (1990). Electrical stimulation of the cerebral cortex in humans. In A. A. Boulton, G. B. Baker & M. Hiscock (Eds.), *Neuropsychology* (Vol. 17, pp. 203–223). Neuromethods. Clifton, New Jersey: Humana Press.
- Maynard Smith, J. (1987). When learning guides evolution. *Nature*, 329, 761–762.
- McClelland, J. L., McNaughton, B. L. & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3), 419–457.
- McDougall, W. (1923). *Outline of psychology*. London: Methuen.
- McGonigle, B. O. & Chalmers, M. (1996). The ontology of order. In L. Smith (Ed.), *Critical readings on Piaget* (Chap. 14). London: Routledge.
- Metta, G., Fitzpatrick, P. & Natale, L. (2006). YARP: Yet another robot platform. *International Journal on Advanced Robotics Systems*, 3(1), 43–48.
- Minsky, M. & Papert, S. (1969). *Perceptrons*. Cambridge, MA: MIT Press.
- Müller, G. B. (2008). Evo-devo as a discipline. *Evolving Pathways: Key Themes in Evolutionary Developmental Biology*, 5–30.
- Myung, J., Forster, M. R. & Browne, M. W. (2000). Special issue on model selection. *Journal of Mathematical Psychology*, 44(1).

- Nehmzow, U., Smithers, T. & McGonigle, B. O. (1993). Increasing behavioural repertoire in a mobile robot. In *From animals to animats 2: (sab '92)* (pp. 291–297). Cambridge, MA: MIT Press.
- Newell, A. & Simon, H. A. (1972). *Human problem solving*. Prentice-Hall.
- Paenke, I., Sendhoff, B. & Kawecki, T. J. (2007). Influence of plasticity and learning on evolution under directional selection. *The American Naturalist*, 170(2), 47–58.
- Partington, S. J. & Bryson, J. J. (2005). The Behavior Oriented Design of an Unreal Tournament character. In T. Panayiotopoulos, J. Gratch, R. Aylett, D. Ballin, P. Olivier & T. Rist (Eds.), *The fifth international working conference on intelligent virtual agents* (pp. 466–477). Kos, Greece: Springer.
- Paul, C. (2004). Morphology and computation. In *From animals to animats 8 (SAB '04)* (pp. 33–38). Los Angeles, CA.
- Pfeifer, R. (Ed.). (1999). *Dynamics, morphology, and materials in the emergence of cognition*. Springer.
- Piaget, J. (1954). *The construction of reality in the child*. New York: Basic Books.
- Port, R. F. & van Gelder, T. (Eds.). (1995). *Mind as motion: explorations in the dynamics of cognition*. Cambridge, MA: MIT Press.
- Prescott, T. J., Bryson, J. J. & Seth, A. K. (2007). Modelling natural action selection: An introduction to the theme issue. *Philosophical Transactions of the Royal Society, B — Biology*, 362(1485), 1521–1529.
- Redgrave, P., Prescott, T. J. & Gurney, K. (1999). The basal ganglia: a vertebrate solution to the selection problem? *Neuroscience*, 89, 1009–1023.
- Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition*, 7, 17–42.
- Riesenhuber, M. & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2, 1019–1025.
- Rogers, T. T. & McClelland, J. L. (2004). *Semantic cognition: a parallel distributed processing approach*. Cambridge, MA: MIT Press.
- Rutte, C. & Taborsky, M. (2007). The influence of social experience on cooperative behaviour of rats (*rattus norvegicus*): Direct versus generalized reciprocity. *Behavioral Ecology and Sociobiology*, 62(4), 499–505.
- Salthouse, T. A. (1986). Perceptual, cognitive, and motoric aspects of transcription typing. *Psychological Bulletin*.
- Seth, A. K. (2007). The ecology of action selection: Insights from artificial life. *Philosophical Transactions of the Royal Society, B — Biology*, 362(1485), 1545–1558.
- Shanahan, M. P. (2005). Global access, embodiment, and the conscious subject. *Journal of Consciousness Studies*, 12(12), 46–66.
- Skinner, B. F. (1935). The generic nature of the concepts of stimulus and response. *Journal of General Psychology*, 12, 40–65.
- Spelke, E. S. (2003). What makes us smart? Core knowledge and natural language. In D. Gentner & S. Goldin-Meadow (Eds.), *Advances in the investigation of language and thought*. Cambridge, MA: MIT Press.
- Sperber, D. & Hirschfeld, L. (2006). Culture and modularity. In P. Carruthers, S. Laurence & S. Stich (Eds.), *The innate mind: culture and cognition* (Vol. 2, pp. 149–164). Oxford University Press.
- Sterelny, K. (2007). *Dawkins versus Gould: Survival of the fittest*. Cambridge, UK: Icon.
- Teyler, T. J. & Discenna, P. (1986). The hippocampal memory indexing theory. *Behavioral Neuroscience*, 100, 147–154.
- Thelen, E. & Smith, L. B. (1994). *A dynamical systems approach to development of cognition and action*. Cambridge, MA: MIT Press.
- Thórisson, K. R., Pennock, C., List, T. & DiPirro, J. (2004). Artificial intelligence in computer graphics: A constructionist approach. *ACM SIGGRAPH Computer Graphics*, 38(1), 26–30.
- Tinbergen, N. (1951). *The study of instinct*. Oxford: Clarendon Press.

- Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society, series 2*, 42(1936–37), 230–265.
- van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences*, 21(5), 616–665.
- Vereijken, B. & Whiting, H. T. A. (1998). Hoist by their own petard: the constraints of hierarchical models. *Brain and Behavioral Sciences*, 21(5), 695.
- West, S. A., Griffin, A. S. & Gardner, A. (2007). Evolutionary explanations for cooperation. *Current Biology*, 17, R661–R672.
- Winslow, B. B., Takimoto-Kimura, R. & Burke, A. C. (2007). Global patterning of the vertebrate mesoderm. *Developmental Dynamics*, 236(9), 2371–81.
- Wolpert, D. H. (1996). The existence of *a priori* distinctions between learning algorithms. *Neural Computation*, 8(7), 1391–1420.