Chapter 25

Evolutionary robotics

Dylan Evans and Walter de Back

Introduction

Robotics has been heavily influenced by cognitive psychology and evolutionary biology, but the child prodigy of these two disciplines, evolutionary psychology, has not yet had such a large impact on the field. Here we describe how evolutionary psychology may come to have a greater impact on robotics in the near future. First, however, we will briefly outline the role that the two parent disciplines of evolutionary psychology have played in the development of robotics, and describe the few lessons that have been drawn so far by roboticists from evolutionary psychology itself.

Cognitive psychology

The cognitive revolution swept through psychology in the 1960s, displacing the behaviourist paradigm that had held sway since the 1920s. Its origins, however, lie in the 1950s. If one day had to be singled out as the birthday of cognitive science, it is surely 11 September 1956. It was on that day that three seminal papers were presented at a historic meeting at the Massachusetts Institute of Technology (MIT). Allen Newell and Herbert Simon spoke about a 'logic theory machine', inaugurating the modern discipline of Artificial Intelligence (Newell and Simon 1956). Noam Chomsky described 'three models for the description of language' in a paper that has been described as marking the birth of modern linguistics (Dennett 1995, p. 384; Chomsky 1956). Finally George Miller presented a paper about short-term memory that is now recognized as one of the foundational papers of cognitive psychology (Miller 1956).

It is somewhat harder to date the origins of robotics, but the pioneering work of Grey Walter and other figures in the cybernetics movement in the late 1940s has a good claim to be considered as laying the foundations for the discipline as we know it today. The mechanical tortoises built by Walter are noticeably different from the robots after the rise of cognitive science a decade later (Walter 1951). For one thing, Walter made no attempt to build a 'mind' out of various functional components such as memory, perceptual processing, and so on. Rather, his emphasis was on the behaviour of the robot as a whole, and on seeking the simplest internal structure that would enable complex behaviour to emerge when the robot was placed in the right kind of environment. This approach may have been largely due to the fact that the integrated circuit had not yet been invented, which meant that Walter had to work entirely with conventional analogue electronic components, large numbers of which would have been required to build components with a distinct psychological function.

The invention of integrated circuits, in which large numbers of components were packed onto a small silicon chip, gave roboticists the means, and the simultaneous rise of cognitive science, the motive, to pay more attention to the internal computational structure of their machines. An early example of the new generation of 'cognitive robots' was Shakey. Developed at the Artificial Intelligence Center at Stanford Research Institute between 1966 and 1972, Shakey had a complex cognitive architecture in which distinct functions such as perception, planning, and natural language processing were implemented by separate programs, which reflected the emphasis of cognitive psychology on the functional decomposition of mental processes.

Evolutionary biology

Biorobotics and evolutionary robotics

By the late 1980s, the cognitive approach to robotics was coming under increasing criticism from Rodney Brooks and his colleagues at MIT. Brooks argued that reasoning and planning via symbol systems, as in Shakey, was often unnecessary, since similar behaviours could be achieved more simply and economically by building robots out of distinct modules, each directly linking sensors to actuators. Unlike cognitive robots, in these 'behaviour-based' robots there was no central processing unit through which all information would be routed. Instead, complex behaviour emerged from the interactions between largely autonomous modules organized into a hierarchy where higher levels could inhibit lower ones.

Brooks and his colleagues were inspired directly by ideas from evolutionary biology. He drew a parallel between his approach to designing robots incrementally, and the stepwise manner by which natural selection builds complexity in nature:

> The advantage of this approach is that it gives an incremental path from very simple systems to complex autonomous intelligent systems. At each step of the way, it is only necessary to build one small piece, and interface it to an existing, working, complete intelligence.
>
> (Brooks 1991, p. 403)

Others have criticized the analogy between the construction of behaviour-based robots and the evolution of organisms. Paul Griffiths (1999, p. 51) is surely right when he states that it is 'implausible that our brains evolved by adding separate mechanisms subserving new functions'. The mind may be massively domain-specific as many evolutionary psychologists contend, but if it is, the various mechanisms surely evolved in parallel, just like the organs of the body.

Brooks also likes to compare the development of artificial intelligence with the history of life on earth, suggesting that it would not be possible to build robots with human-level intelligence without first endowing them with less complex capacities such as locomotion. From this perspective, the construction of Deep Blue, a chess-playing computer which won a six-game match against world champion Garry Kasparov in 1997, was an ambiguous achievement. Since many people find chess cognitively-challenging, it was perhaps natural for the pioneers of artificial intelligence to regard chess-playing as a legitimate goal for computers, but Brooks viewed such projects as dead ends. While IBM was building Deep Blue, other researchers were struggling to build robots that could walk over anything except a perfectly flat surface without falling over. Brooks argued that the flexibility of human intelligence would only be replicated in robots that had first acquired a bedrock of capacities, such as locomotion, that had evolved much earlier than abstract reasoning.

The robots built by Brooks were inspired by general principles of evolutionary biology, and sometimes by generic features of large classes of organisms such as hexapod locomotion, but were not generally intended to replicate particular species. During the 1990s, however, researchers began to use robots to test detailed hypotheses about the biological mechanisms of individual species. Barbara Webb, for example, developed a spiking neuron network to model the process by which crickets processed acoustic signals, and mounted the circuit on a mobile robot to test her hypotheses about how female crickets locate males by detecting their songs (Webb 1995). Other roboticists have used a similar approach to test hypotheses about whisking in rats and navigation in ants, giving rise to a dynamic subfield known as 'biorobotics' (Webb 2001).

A related area of research, known as 'evolutionary robotics' (Nolfi and Floreano 2000), uses genetic algorithms and other types of evolutionary computation to develop artificial neural networks (and sometimes robot morphology). Most experiments in evolutionary robotics follow the same basic format. First, a population of artificial chromosomes that code for the control system and/or the morphology of a robot are randomly generated. Next, these chromosomes are decoded and the resulting robot (which may be physical or simulated) is set free to act in a given environment while its performance on various tasks is automatically evaluated according to a predefined criterion known as the fitness function. The fittest robots are then allowed to reproduce by generating copies of their genotypes, with the addition of changes introduced by some genetic operators such as mutation and recombination. This process is then repeated for several generations, allowing selection to enhance the average fitness of the population in the same way as it has done so effectively in the history of life as we know it.

Most current research in biorobotics and evolutionary robotics tends to fall into one of the following categories:

Navigation

A capacity for navigation is one of the most basic requirements for an autonomous mobile robot, so it is unsurprising that this remains a focus of evolutionary robotics research. Typical tasks that robots are set include obstacle avoidance and navigating toward a target area. Since most animal navigation is visually guided, this field of research includes a great deal of work on visual sensing, including such areas as object recognition and discrimination, though many robots still use infrared sensors or sonar instead of, or in addition to, a camera.

Competitive co-evolution

Most experiments in the evolution of robot navigation involve testing each member of each generation, one-by-one, in a fixed physical environment. Experiments in co-evolution, by contrast, involve a continually changing dynamic environment consisting of other agents. Biologists have long speculated that such an environment may enhance the adaptive power of natural selection. For example, Dawkins and Krebs (1979) argued that competing populations may reciprocally drive one another to increasing levels of behavioural complexity by producing an evolutionary arms race. Researchers in artificial life began to develop virtual simulations of such arms races in the early 1990s (Miller and Cliff 1994), and more recently, researchers in evolutionary robotics have developed fully-embodied physical models of predator–prey dynamics.

Cooperative collective robots

Experiments in collective robotics need not be entirely competitive. A thriving strand of research studies the evolution of cooperation among teams of mobile robots. Such experiments often involve modelling the behaviour of social insects (e.g. ants, bees) which display sophisticated collective intelligence. This swarm intelligence (Bonabeau et al. 1999) emerges from the networks of interactions among agents that are individually quite simple. Research in this area includes work on social signalling; Luc Steels (1998), in a pioneering series of experiments, used robots to explore various hypotheses about the origins of language.

Evolvable hardware

Most research in evolutionary robotics takes the robot body as a fixed parameter, focusing exclusively on evolving robot control systems. However, a few projects have applied the evolutionary process to robot bodies. The most striking experiment so far in this field involved the evolution of simple locomotive systems composed of bars and actuators, and then used rapid-prototyping technology to produce the multi-linked structures, so that the only human input needed was the final attachment of snap-on motors (Lipson and Pollack 2000). Another approach in evolvable hardware is evolving robot controllers in reconfigurable electronic circuits such as field-programmable gate arrays (FPGAs). These are argued to be analogue dynamical continuous-time systems and thereby avoid the constraints of discrete digital design (Thompson 1997).

The explicit discussion of evolvable hardware draws attention to an aspect of evolutionary robotics that may not be apparent from the discussion so far: the fact that experiments can be done by means of computer simulations as well as with real, physical robots. The relative advantages and disadvantages of each approach are the subject of some debate in evolutionary robotics. Some argue that physical robots are better models since they incorporate real problems in coping with physical forces like sensory noise, energy consumption, damage, and inertia. However, experiments involving physical robots are very time-consuming. Simulated robots avoid this problem, but also circumvent the physical problems that may well be essential to the structure of natural minds. Another important problem with physical robots is the inability to evolve bodily structures (evolvable hardware as described above is a rare exception), which is essential for the modelling of co-evolution of body and brain. This is so far only possible in simulation. Hybrid approaches perform the most time-consuming periods of evolution in simulation, and subsequently evolve the controllers further in physical robots.

Evolutionary psychology

The previous two sections have provided a brief summary of the huge impact that the two parent disciplines of evolutionary psychology—cognitive psychology and evolutionary biology—have each had on the development of robotics. Given this legacy, it is somewhat surprising that evolutionary psychology has itself had little influence so far on robot research. In this section we discuss some ways in which evolutionary psychology may be applied to the development of robots in the near future.

Synthetic evolutionary psychology

The methods used by evolutionary psychologists so far may be described as analytic, in the sense that they start with a real system (the minds of modern humans and/or of various ancestral species) and attempt to collect data about this system that might permit us to infer the internal structure of this system. There is nothing inherently wrong with analytic methods—indeed, they are the backbone of most modern science—but researchers are increasingly aware of their drawbacks. One important drawback is that the difficulty of analysing a system grows exponentially with the complexity of the system. Since minds, especially the minds of advanced primates such as our own and those of our recent ancestors, are notoriously complex systems, it follows that all analytic methods are very hard to apply to the study of the human mind in a way that all researchers agree upon. One has only to look at the voluminous polemical literature that has arisen in response to the version of the Wason selection task originally described by Cosmides and Tooby, to realize how difficult it is to derive uncontroversial conclusions from analytic methods in evolutionary psychology (Cosmides and Tooby 1992).

Valentino Braitenberg (1984) has proposed, in respect of such difficulties, that when it comes to complex systems it is often easier to discover their internal structure by synthetic methods. In other words, if we wish to discover how some system works, it is often easier to do so by building successively more complex models, rather than by attempting to infer the mechanism from mere observation:

> It is pleasurable and easy to create little machines that do certain tricks. It is also quite easy to observe the full repertoire and behaviour of these machines even if it goes beyond what we had originally planned, as it often does. But it is much more difficult to start from the outside and to try to guess internal structure just from the observation of behaviour.
>
> (Braitenberg 1984, p. 20)

Braitenberg refers to this generalization as the law of uphill analysis and downhill synthesis. But this way of putting things is misleading to the extent that it suggests that the researcher is faced with a choice between analytic and synthetic methods. In reality, analytic and synthetic methods are not alternatives, but complementary aspects of a dialectic that involves moving back and forth between the analysis of empirical data and the construction of simple models of underlying mechanisms.

Synthetic methods involve building models of the system under investigation and then observing their behaviour. The more closely the behaviour of the model corresponds to the behaviour of the target system, the more confident we can be that the internal structure of the model corresponds to the internal structure of the target. Because we have built the model ourselves, its internal structure is transparent, and need not be inferred by analysis.

With a few notable exceptions (such as Geoffrey Miller, Gerd Gigerenzer, and Douglas Kenrick), most evolutionary psychologists have not attempted to translate their hypotheses about human mental structure into working machines. This might seem to exclude evolutionary psychology from cognitive science, a core feature of which is often taken to be the design-based approach. However, this conclusion is too quick, for emphasis on design does not imply that all cognitive scientists must take an active part in building artificial minds. It simply requires that cognitive scientists propose models of the mind that are computational enough to permit computer programs to be readily designed on the basis of the models. Many of the domain-specific mechanisms proposed by evolutionary psychologists take such a form; they are not specified in terms of any programming language, but they are often spelled out in a form that would be relatively easy to convert into a computer program. The models proposed by evolutionary psychologists thus count as fully cognitive, and evolutionary psychology is firmly within the fold of cognitive science.

Even so, it seems a shame that evolutionary psychologists have not taken more interest in translating their models into real machines. The tools of artificial intelligence and computational modelling might well offer them ways of testing their hypotheses and thus enable them to answer the common charge of telling 'just-so stories'. Critics of evolutionary psychology frequently dismiss it on the grounds that it promulgates untestable theories. Evolutionary psychologists acknowledge that there are methodological difficulties posed by investigating the history of the mind, but point out that most of these difficulties are not particular to their discipline. Most of them are common problems faced by all those who wish to investigate evolutionary hypotheses, so to be consistent the critics should also dismiss the whole of evolutionary biology. Such general arguments, however, would be strengthened if evolutionary psychologists could also point to experimental ways of testing their hypotheses. Robotics could supply evolutionary psychology with just such experimental techniques.

Computer simulations of evolutionary processes are already common. In one early example, Thomas Ray (1992) designed a virtual world called Tierra and populated it with a simple digital organism. This was a simple self-replicating program which occasionally made mistakes in the copying process. These 'mutations' led to an increasingly diverse population of digital organisms. Competition for limited memory space on the computer's hard disk ensured that there was differential survival. The conditions for natural selection were therefore all in place, and Ray was able to observe numerous cases of digital evolution complete with virtual viruses, parasite resistance, and other surprisingly 'natural' features.

Tierra is only a virtual world, and thus subject to the criticisms of roboticists like Rodney Brooks, who argue that it is all too easy in such simulations to make some crucial but unnoticed simplification that renders the simulation invalid. In order to avoid this potential danger, Brooks recommends that cognitive scientists work with actual physical robots. By using the techniques of evolutionary robotics, evolutionary psychologists could observe artificial evolution in the real world, giving rise to a genuine lineage of robots evolving by natural selection. Perhaps these methods can provide a way for evolutionary psychologists to test their hypotheses about mental evolution.

As an example of such an approach, den Dulk et al. (2003) studied the evolution of dual-route dynamics for affective processing. In particular, they used the standard methods of evolutionary robotics to examine the evolutionary justification given by LeDoux (1998) for his dual-route model of fear-processing. LeDoux has found evidence that, in many mammals, fear is processed simultaneously by two neural pathways, one subcortical and the other largely cortical. The subcortical route is faster but generates many false positives, while the cortical route is slower but more accurate. LeDoux argues that this dual-route mechanism evolved by natural selection because it allowed animals to get the best of both worlds by escaping quickly when necessary, but not wasting too much effort on false alarms. By allowing agents to evolve in a simple environment consisting of predators and food, den Dulk et al. (2003) found that agents did indeed evolve a dual-route mechanism similar to that proposed by LeDoux, but only when certain conditions were met: the food and the predator had to be relatively hard to distinguish, and information must take significantly longer to propagate via the cortical route than via the subcortical route.

A programme of research in synthetic evolutionary psychology might use much of the same tools as those used by mainstream research in evolutionary robotics, but the issues for investigation would be substantially different. As we have already seen, evolutionary robotics is generally concerned with navigation, co-evolution, cooperation, and evolvable hardware. Evolutionary psychologists would instead be more interested in using robots to explore psychological questions such as the history and structure of the various mental modules that comprise the human mind.

A representative sample of the kinds of problems that are thought by evolutionary psychologists to have led to the evolution of mental modules include: 1) finding food and discriminating between nutritious substances and toxins; 2) detecting and avoiding predators; 3) eliminating infectious agents from the body; 4) establishing and maintaining strategic alliances; 5) communicating with conspecifics; 6) finding and keeping mates; and 7) rearing offspring.

Experiments in evolutionary robotics have already begun to explore the first two problems in this list, and have touched on the problem of communication, but have left the other problems virtually unexplored. This is partly due to technological limitations, but such considerations apply only to conducting experiments with real physical robots. Computer simulations of evolving robot populations could explore the other areas without much difficulty. There is great scope, then, for a broad research programme to explore all of these problems in a systematic way, perhaps by a graded approach that tackles each problem in the order in which they were faced by our ancestors. Such an incremental approach might also throw light on the way in which prior adaptations may be co-opted by natural selection as the basis for solutions to later problems.

Open-ended evolution

A well-known and important difference between artificial and natural evolution is that the former is goal-directed, while the latter is open-ended. Artificial evolution is a technique originally devised to optimize certain parameters. Natural evolution, in contrast, does not converge to a single solution and then stop: it is open-ended and continuous. There are no optimal solutions in nature because the problems, arising from the biological context which includes many co-evolving creatures, are constantly changing. Evolutionary psychologists are interested in the mechanisms that result from natural, and therefore open-ended, evolution. A synthetic approach to evolutionary psychology should attempt to replicate this process. Over the years, various technical procedures have been proposed that partly overcome the major differences between artificial and natural evolution, such as massive co-evolution, variable genotype length, and complex genotype/phenotype mapping schemes.

The goal-directed nature of artificial evolution is mainly due to the use of fitness functions, which are mathematical formulae used to calculate the relative fitness of an individual. Selection mechanisms operate on the basis of the relative differences between fitness values: better individuals are selected for reproduction. After many iterations, evolution thus selects the individuals that optimize the components of the fitness function. In evolutionary robotics, the task of robot is thus implicitly coded in the fitness function (although the behavioural components are not described). Floreano and Urzelai (2000) propose the fitness space as a framework in which fitness functions can be positioned consisting of three dimensions: 1) *Functional/behavioural*: is fitness dependent on the internal functioning of the controller or the behavioural manifestation? 2) *Explicit/implicit*: are there many or very few variables included in the fitness function? 3) *External/internal*: are the variables in the fitness function only available to the fitness evaluator or to the evolving robot as well? According to this framework, the key to approximate open-ended evolution as found in nature is to make a fitness function as behavioural, implicit, and internal as possible.

Alternative evolutionary schemes

In evolutionary theory, fitness is defined in terms of relative reproductive success. This is incompatible with standard artificial evolution on pain of tautology: selection for reproduction cannot be based on relative fitness if this is itself based on reproductive success, because it implies that reproduction already took place. A more biologically plausible evolutionary scheme for experiments in synthetic evolutionary psychology would therefore be one in which selection is not based on explicit fitness values. Instead, selection should emerge from the interaction of the agents with their environment. Fitness values should be used solely for the purpose of cross-comparison and other analyses of the experimental results. This will have impact on the overall set-up of an experiment, which will render it more biologically plausible, without making it necessarily more detailed or complex.

An experiment of this kind should have populations of robots roaming the environment at the same time, in contrast to standard evolutionary robotics where individual controllers are often tested one at a time on the same physical robot. Having simultaneous individuals introduces problems of survival and reproduction if resources are limited. The adaptations to cope with such problems are precisely the ones in which we are interested in evolutionary psychology. Furthermore, having simultaneous populations of potential mates implies continuous reproduction. This results in a gradual generational change, in contrast to many evolutionary robotics experiments where generations are discrete in the sense that they are separated by a selection operation. Gradual generational change and simultaneous populations of individuals means having different (grand)parents and offspring in the same environment, which together with the problem of survival (and growth), introduces the problem of parental care. This would also allow adaptations for social transmission and cultural evolution to evolve. Even a few steps closer to a biologically plausible context should allow sexual selection to influence morphology, having the

OUP UNCORRECTED PROOF-FPP,18/08/2011, CENVEO

environment include predation, or even a full-fledged co-evolving food chain. The possibilities of this kind of experiment are of course limited by the constraints imposed by physical design (e.g. appropriate sensors and effectors), adaptability of robot controllers (e.g. sufficiently complex neural networks), mechanical issues (e.g. energetic autonomy), and many other technical limitations.

Experimental design in synthetic evolutionary psychology

Incorporating the alternative evolutionary scheme just described would have a substantial impact on experimental design. Because fitness functions are omitted in the experimental set-up, there are no means to formulate a specific task, which is the first step in designing a standard experiment in evolutionary robotics. Research in synthetic evolutionary psychology should proceed differently. Although specifying an entire experimental framework of synthetic evolutionary psychology is beyond the scope of this chapter, and involves long and thorough experimentation, a few important aspects are noted here.

A hypothesis about the emergence or structure of a particular mental module or behavioural trait is generally based on the following aspects: (changing) environmental conditions, evolutionary forces, and already existing mental structures. That is, an adaptation emerges from new or changed situations in the individual's environment; it benefits the organism's ability to cope with the evolutionary forces; and it is based on existing mental structures or neural substrates that facilitate the emergence of new abilities. Verifying such a hypothesis involves synthesizing these aspects in the experiment and analysing evolving individuals for emerging adaptations in coherence with changing environments.

Changing environmental conditions have played a central role in the evolution of all species. Either it gradually forces adaptation in a species by slow environmental change such as diminishing food resources, or it radically forces selection by historical natural accident like a meteor impact or epidemic. Researchers disagree about the relative importance of these two kinds of environmental pressures, but everyone agrees that both have played a part in the evolution of the human mind. It is relatively easy to synthesize both kinds of environmental changes in conditions in synthetic evolutionary psychological experiments. This can be used to test the robustness of certain adaptations, to cause the emergence of new adaptations, and to provide insight into the trade-off between the importance of robust processes and actual sequences in explaining the history of the mind. Experimenters could slowly increase or decrease the presence of predators, food resources, and mates. Historical accidents can be simulated through sudden changes in availability of resources or removal of individuals.

The existing mental structures that facilitate the emergence of a certain adaptation can be synthesized in two ways that resemble the different goals in evolutionary psychology. If one is particularly interested in exploring the historical trajectory of the mind, the mental structures that facilitate certain adaptations are themselves the results of earlier evolutionary experiments. That is, a graded or incremental approach is taken to evolve increasingly complex robot controllers. This is achieved by increasing the difficulty to survive and reproduce by adding more and more environmental complexity. On the other hand, if one is more interested in the structure of the mind, a somewhat more pragmatic approach can be adopted. In these experiments, the existence of particular structures can be assumed. These structures can, for example, be synthesized by means of neural controllers that are trained by using supervised feedback, without concern for the evolutionary plausibility of their existence. Although this undermines the historical account of adaptation, it can prove to be a fruitful and time-saving approach to investigate the emergence of higher-level adaptations.

25-Roberts-Ch25.indd 421 8/16/2011 4:30:46 PM

Once these three conditions have been specified and synthesized, the individual creatures must be designed with the appropriate sensors, motors, and neural controllers. Experimenters subsequently look for evidence for adaptations in changing behaviour, examine internal functioning of the controller and robustness to changing conditions, analyse fitness changes, make comparisons with naturally evolved creatures, or perform lesions and look for dissociations. Plausible stories about the historical evolution of situated, embodied minds can then be supported or eliminated by hard evidence.

Synthetic evolutionary psychological research has the potential to yield new, or confirm existing, hypotheses about the evolution of specific adaptations. Furthermore, it can provide insight about the importance of the different variables in the evolution of natural minds. It can provide us with increasingly complex machines with needs comparable with those of living creatures. And perhaps these machines will exhibit cognitive abilities, evolved out of the most primal forces in nature, comparable with ours.

The application of evolutionary psychology to human–robot interaction

Some might argue that the research programme we have outlined in the previous section is out of place in a book about 'applied evolutionary psychology', on the grounds that this programme is more about the application of robotics to evolutionary psychology than vice versa. We would disagree with such a claim, since our suggestions about open-ended evolution and alternative evolutionary schemes, which are directly inspired by evolutionary theory, would clearly modify the existing experimental protocols in evolutionary robotics. Furthermore, some roboticists are already carrying out experiments along the lines we have described, and have been directly inspired by ideas from evolutionary psychology in doing so. For example, Alan Winfield and Frances Griffiths are building an artificial society of robots in order to model the mechanisms by which a group of organisms might make the transition from the merely social to the cultural (Winfield and Griffiths 2010). The project is inspired by a memetic theory of culture (Blackmore 1999), itself inspired by Richard Dawkins' speculations about non-genetic replicators at the end of *The Selfish Gene* (Dawkins 1976).

Nevertheless, if we want a less ambiguous example of how robotics might benefit from ideas drawn from evolutionary psychology, we should turn to the study of human–robot interaction. When robots are designed specifically to have rich psychological interactions with humans, as in pedagogical robots and artificial companions (Wilks 2010), rather than merely physical interactions such as surgical robots and military robots, the designers ignore evolutionary psychology at their peril, or must laboriously rediscover for themselves what they could have learned more quickly from evolutionary psychologists. A case in point is the so-called 'uncanny valley hypothesis' first advanced by the Japanese roboticist Masahiro Mori (1970). He proposed that the relationship between a robot's degree of resemblance to a human and a human observer's emotional response was highly non-linear. Robots lacking humanoid features would, he speculated, produce a neutral response in observers, but as robots become progressively more humanlike, we would experience a more positive emotional reaction to them. However, at the point just before it became impossible to distinguish between robots and humans, the robot would evoke in us a sense of revulsion and horror. In reference to the shape of the graph he produced to illustrate his hypothesis (see Figure 25.1), Mori dubbed this zone 'the valley of the uncanny'.

Whether Mori's use of the word 'uncanny' was intended to evoke Freud's earlier essay on this particular feeling is unclear, but in that essay Freud (1955 [1919]) specifically refers to 'ingeniously constructed dolls and automata' as prime examples of things that seem spooky. Citing a

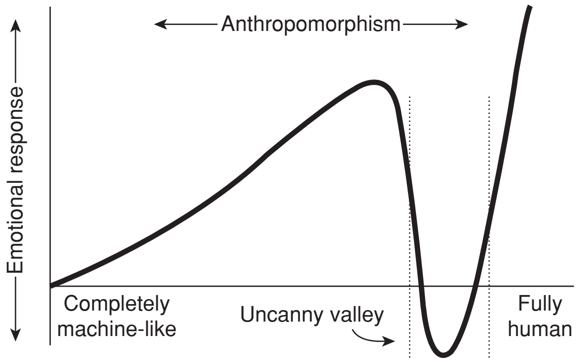

Fig. 25.1 The valley of the uncanny.

1906 essay by Ernst Jentsch, Freud notes that 'in telling a story one of the most successful devices for easily creating uncanny effects is to leave the reader in uncertainty whether a particular figure in the story is a human being or an automaton'. (However, Freud's explanation for this phenomenon, which traces it to the fear of castration, is bizarre and unconvincing.)

Various evolutionary explanations for the uncanny valley have been proposed, all of which are more convincing than Freud's, and which are not mutually inconsistent. One explanation focuses on the psychological mechanisms for mate selection. According to this view, automatic, stimulus-driven appraisals of uncanny stimuli elicit aversion by activating an evolved cognitive mechanism for rejecting potential mates with low fertility, poor hormonal health, or ineffective immune systems, based on visible features of the face and body that are predictive of those traits (Green et al. 2008). Another explanation focuses on pathogen avoidance: uncanny stimuli may activate a cognitive mechanism that originally evolved to motivate avoidance of potential sources of pathogens by eliciting a disgust response. Thus, the visual anomalies of android robots have the same effect as those of corpses and visibly diseased individuals: the elicitation of alarm and revulsion (Green et al. 2008).

Those who build robots intended for rich psychological interactions with humans could also benefit from evolutionary psychology when writing the software for such machines. For example, evolutionary psychology has focused attention, in a way that other branches of psychology have not, on the cognitive mechanisms that underlie reciprocity and fairness. Those who wish to build robot companions might glean much from this literature, perhaps yielding some counter-intuitive guidelines for robot design. For example, it may be that a robot companion who does not punish its owner (in some appropriate way) when the owner treats it unfairly is less engaging than one who does (Evans 2010).

Conclusion

During its relatively brief history, evolutionary psychology has used many different methods, from experimental manipulation of human behaviour in the laboratory to observation of indigenous peoples and analysis of archaeological data. All these methods may be called analytic, in the sense that they collect data about already-existing systems and then analyse them. We have argued that evolutionary psychologists could benefit from extending their methodological repertoire to include synthetic methods, which involve constructing artificial systems. Such artificial systems can provide useful models of evolved minds and evolutionary histories that might provide evolutionary psychologists with additional data to test hypotheses about mental structure and evolutionary trajectories. One kind of synthetic method in which evolutionary psychologists have so far shown little interest, is evolutionary robotics. We have argued that, by ignoring this field, evolutionary psychologists are missing out on a valuable research tool, and we have sketched out a research programme involving the use of robots to test evolutionary psychological hypotheses.

We have referred to some initial work that uses robots to test hypotheses about human evolution. At the moment, however, this work is too rare and patchy to warrant description as a full-fledged research programme. Our manifesto for a new field of synthetic evolutionary psychology remains, for the moment at least, more of a vision than a reality. Time will tell whether this vision is embraced by enough researchers to make it a permanent and fruitful addition to the evolutionary psychologist's repertoire of methodologies.

Acknowledgements

We are grateful to Geoffrey Miller, Verena Hafner, Lars Zwanepol, Tijn van der Zant and Rens Kortmann for their comments. Dylan Evans' work was supported by a Platform Grant from the Engineering and Physical Sciences Research Council, EPSRC GR/M97503.

References

Blackmore, S. (1999). *The meme machine*. Oxford University Press, Oxford.

Bonabeau, E., Dorigo, M., and Theraulaz, G. (1999). *Swarm intelligence: from natural to artificial systems*. Oxford University Press, New York.

Braitenberg, V. (1984). *Vehicles: experiments in synthetic psychology*. MIT Press, Cambridge, MA.

Brooks, R.A. (1991). Intelligence without representation. In: J. Haugeland (ed.), *Mind design II: philosophy, psychology, artificial intelligence*, pp. 395–420. MIT Press, Cambridge, MA.

Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information Theory*, **2**, 113–24.

Cosmides, L. and Tooby, J. (1992). Cognitive adaptations for social exchange. In: J. Barkow, L. Cosmides, and J. Tooby (eds), *The adapted mind: evolutionary psychology and the generation of culture*, pp. 163–228. Oxford University Press, Oxford.

Dawkins, R. (1976). *The selfish gene*. Oxford University Press, Oxford.

Dawkins, R. and Krebs, J.R. (1979). Arms races between and within species. *Proceedings of the Royal Society B*, **205**, 489–511.

den Dulk, P., Heerebout, B.T., and Phaf, R.H. (2003). A computational study into the evolution of dual-route dynamics for affective processing. *Journal of Cognitive Neuroscience*, **15**, 194–208.

Dennett, D. (1995). *Darwin's dangerous idea: evolution and the meanings of life*. Allen Lane, London.

Evans, D. (2010). Wanting the impossible: the dilemma at the heart of intimate human-robot relationships. In: Y. Wilks (ed.), *Close engagements with artificial companions. Key social, psychological, ethical and design issues*, pp. 75–87. John Benjamins, Amsterdam.

Floreano, D. and Urzelai, J. (2000). Evolutionary robots with on-line self-organization and behavioral fitness. *Neural Networks*, **13**, 431–43.

Freud, S. (1955 [1919]). The 'Uncanny'. In: J. Strachey (ed.), *The standard edition of the complete psychological works of Sigmund Freud*, Vol. XVII, pp. 217–56. Hogarth Press, London.

Green, R.D., MacDorman, K.F., Chin-Chang, H., and Vasudevan, S. (2008). Sensitivity to the proportions of faces that vary in human likeness. *Computers in Human Behavior*, **24**, 2456–74.

Griffiths, P. (1999). Author's response. *Metascience*, **8**, 49–62.

LeDoux, J. (1998). *The emotional brain: the mysterious underpinnings of emotional life*. Weidenfeld and Nicholson, London.

Lipson, H. and Pollack, J. (2000). Automatic design and manufacture of robotic lifeforms. *Nature*, **406**, 974–8.

Miller, G.A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, **63**, 81–97.

Miller, G.F. and Cliff, D. (1994). Protean behaviour in dynamic games: arguments for the co-evolution of pursuit-evasion tactics in simulated robots. In: D. Cliff, P. Husbands, J. A. Meyer, and S. Wilson (eds), *From animals to animats 3: Proceedings of the Third International Conference on Simulation of Adaptive Behaviour*, pp. 411–20. MIT Press, Cambridge, MA.

Mori, M. (1970). Bukimi no tani. [The uncanny valley] (translated by K.F. MacDorman and T. Minato). *Energy*, **7**, 33–5.

Newell, A. and Simon, H.A. (1956). The logic theory machine- a complex information processing system. *IRE Transactions on Information Theory*, **2**, 61–79.

Nolfi, S. and Floreano, D. (2000). *Evolutionary robotics: the biology, intelligence and technology of self-organizing machines*. MIT Press, Cambridge, MA.

Ray, T.S. (1992). An approach to the synthesis of life. In: M.A. Boden (ed.), *The philosophy of artificial life*, pp. 111–45. Oxford University Press, Oxford.

Steels, L. (1998). The origins of syntax in visually grounded robotic agents. *Artificial Intelligence*, **103**, 133–56.

Thompson, A. (1997). Artificial evolution in the physical world. In: T. Gomi. (ed.), *Evolutionary robotics: from intelligent robots to artificial life*, pp. 101–25. AAI Books, Ottawa, ON.

Walter, W.G. (1951). A machine that learns. *Scientific American*, **185**, 60–3.

Webb, B. (1995). Using robots to model animals: a cricket test. *Robotics and Autonomous Systems*, **16**, 117–34.

Webb, B. (2001). Can robots make good models of biological behaviour? *Behavioral and Brain Sciences*, **24**, 1033–50.

Wilks, Y. (ed.) (2010). Close engagements with artificial companions. *Key social, psychological, ethical and design issues*. John Benjamins, Amsterdam.

Winfield, A.F.T. and Griffiths, F. (2010). Towards the emergence of artificial culture in collective robot systems. In: P. Levi and S. Kernbach (eds), *Symbiotic multi-robot organisms*, pp. 431–9. Springer, Berlin.