Biological Style Evolution of Robot Control Systems

John A. Bullinaria
School of Computer Science
The University of Birmingham
Birmingham, B15 2TT, UK
j.bullinaria@physics.org

Abstract

In attempting to formulate realistic models of the development of the human oculomotor control system, it became clear that evolutionary factors played a crucial role. Even rather coarse simulations of the biological evolutionary process resulted in adaptable control systems considerably more efficient than those I had attempted to build 'by hand'. I this paper I shall describe some of the aspects of these models that are likely to be useful for building robot control systems. In particular, I shall look at the evolution of appropriate innate starting points for learning/adaptation, patterns of learning rates that vary across different system components, individual differences, and learning rates that vary during the system's lifetime.

1   Introduction

For several years I have been involved in formulating increasingly sophisticated models of the development of human oculomotor control with view to better understanding the problems that may occur under normal conditions, and also as a result of using novel viewing devices such as virtual reality head mounted displays (e.g. Bullinaria, Riddell & Rushton, 1999; Bullinaria & Riddell, 2001). These models were rather successful in that they could simulate adult human performance reasonably well, and also the developmental progression towards those abilities, yet they remained lacking in that they still involved assumptions about various innate factors that were difficult to determine empirically (Bullinaria & Riddell, 2000). Particularly important were the innate starting points of the learning process, the advantages of having different learning rates for distinct components of the system, and the relevance of the critical periods of learning that are observed in humans. Some preliminary small scale simulations of human like evolution of these factors (Bullinaria, 2001a,b) suggest that such an approach will not only result in improved models of human systems, but could well result in improved performance for analogous robotic systems.

The idea of using evolutionary techniques for creating robotic systems is already well established in the field of *Evolutionary Robotics* (e.g. Nolfi & Floreano, 2000). Evolutionary algorithms have also already shown much promise for generating artificial neural networks with performance superior to those formulated directly by human researchers. Factors such as network architecture, learning rules and connection weights have all been successfully optimised by evolution (e.g. Yao, 1999). In this paper, however, I would like to concentrate on the effects of three related factors that are crucial for the human system, and yet have perhaps not received adequate attention in the field of robotics. First, the so-called *Nature-Nurture Debate* and the distinctions between properties that are innate in each new individual and those that must be learned from the environment during an individual's lifetime. Second, the range of *Individual Differences* and how these are constrained by evolution and learning. Third, the advantages and disadvantages of the *Critical Periods for Learning* that are often observed in human development. I shall argue that to study these issues properly, it is important that the evolutionary simulations are more closely aligned with biological evolution than is commonly the case.

2   The Baldwin Effect

Clearly we are primarily interested in how evolution by natural selection can improve our system's (robot's) ability to act in and learn from its environment. Inevitably, that interaction between learning and evolution known as the *Baldwin Effect* (Baldwin, 1896; Belew & Mitchell, 1996) will be crucial for understanding the processes involved. For present purposes this interaction occurs in two stages: (1) if a mutation improves the ability of the learning process to acquire better properties, then it will tend to proliferate in the population, and (2) if the learning process has an associated cost, then its results will tend to get incorporated into the genotype and the learned behaviours will become innate. In effect we will have genetic assimilation, without Lamarckian inheritance, of acquired characteristics. However, if the system really does need to retain the ability to learn, for example to adapt to unknown or changing environmental conditions, or to adapt to the system's own maturation (as we generally have in biological systems), then
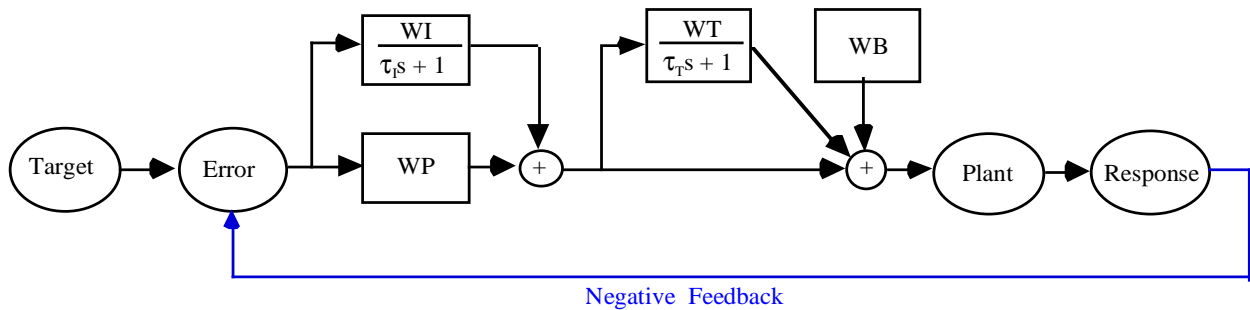
Figure 1: A simplified control model with four learnable weights/parameters: WI, WP, WT, WB.

we will only get partial assimilation. We should still end up with an efficient learning system, but the appropriate innate properties will not generally correspond to the final learned behaviour. This is where the nature-nurture debate comes in. Moreover, if natural selection cannot discriminate between different genotypes (i.e. different innate properties), we will be left with a range of individual differences. I have recently explored these two issues in more detail elsewhere (Bullinaria, 2001a).

3   Variable Plasticity

The third issue I wish to explore is how evolution might lead to more efficient systems by allowing the emergence of variable plasticities (i.e. non-constant learning rates). It is well known that human neural plasticity varies considerably with age, and that there are often critical periods during which learning must take place if the given task is to be mastered successfully (Greenough, Black & Wallace, 1987; Julesz & Kovacs, 1995). The idea of variable neural plasticity is also quite common in the field of artificial neural networks where modellers have found it beneficial to vary their network learning rates during the course of training (Jacobs, 1988). For example, near the end of training it may be useful to decrease the learning rates to minimise the weight variations seen after each sample in online training, or to increase them to speed the saturation of sigmoids as the errors become small. Alternatively, if the performance of a task depends crucially on some lower level of processing, it may be sensible to delay the learning of that task until the lower level processes have fully developed. It is not clear to what extent factors such as these have been responsible for the evolution of the patterns of plasticity found in humans, or if it has been more a matter of minimizing the physical overheads of the plasticity.

4   The Control Model

In this paper I shall present the results of a series of explicit simulations of the evolution of some simple adaptable control systems. The overall aim will be to see which learning strategies evolve naturally, and to explore how

different strategies evolve under different circumstances. These can be expected to inform the issues discussed above, and will hopefully enable us to formulate better adaptable controllers for our robotic systems.

The control system that will form the basis of the current investigation is shown in Figure 1. It is actually a simplified version of the part of the oculomotor control system that focuses and rotates the human eye (Schor et al., 1992), though similar systems can be applied quite generally (Levine, 1996). The input is a natural sequence of target responses and the feedback loop allows the determination of an error signal. This signal then feeds into a standard integral controller (with gain/weight WI) and a standard proportional controller (with gain/weight WP), the outputs of which are added to a constant bias signal (of strength/weight WB) and a leaky integrator tonic signal (of gain/weight WT), and fed into the (leaky integrator) plant to produce the response. The bias provides an appropriate general resting state, while the tonic allows short time-scale adaptation of the resting state during periods of constant demand. In the human eye focusing system, for example, we would have blur being processed to generate signals for the ciliary muscles in the eye appropriate for the distance of the visual target (Schor et al., 1992). The system can equally well be regarded as a traditional control system (Levine, 1996), or as a dynamical network of leaky integrator neurons (Bullinaria & Riddell, 2001).

Simulating the evolution of this system will involve working with a large number of copies of the model, each with their own four adjustable parameters/weights $\mathbf{W}(t) = \{WI(t), WP(t), WT(t), WB(t)\}$ where t is the time/age of the individual model measured in simulated years. These parameters are learned by a simple on-line gradient descent algorithm that minimizes a cost function consisting of response error and smoothness (regularization) components, which would be readily available to the system, for representative sequences of response targets (Bullinaria & Riddell, 2001). Corresponding to the learnable weights, then, each instantiation of the model will have four variable learning rates/plasticities $\mathbf{P}(t) = \{PI(t), PP(t), PT(t), PB(t)\}$. The model will also have various other parameters (the time constants

τ_I and τ_T, plant characteristics, feedback time delay, and so on) which we take to be the same for all instantiations, with values appropriate for human oculomotor control (Schor et al., 1992). Such a system that has evolved/learned a good set of weights will produce appropriate damped responses to arbitrary discontinuous output requirements, and smooth pursuit of arbitrary continuous output changes (Bullinaria & Riddell, 2001).

For the purposes of this paper, I shall assume that all the learning rates in a given model vary with age in the same manner, and that this variation depends only on the genotype (innate parameters) of the individual, and not on the environment that the individual finds itself in. Naturally, it will be important to relax this condition in the future, but this means that we can write $\mathbf{P}(t) = s(t).\mathbf{P}(0)$, where $\mathbf{P}(0)$ are the four initial learning rates, and $s(t)$ is a simple age dependent scaling factor. Clearly, if there is no plasticity variation we will have $s(t) = 1$ for all t. A convenient parameterization is to simply take $s(t)$ to be piecewise linear with parameters $\mathbf{S} = \{s(t) : t = 1, \ldots, N\}$. This extends my earlier study (Bullinaria, 2001b) where I took $s(t)$ to be an exponential function determined by only two evolvable parameters. The part of the model's genotype that varies between individuals thus represents the $8+N$ parameters $\{\mathbf{W}(0), \mathbf{P}(0), \mathbf{S}\}$. There is clearly nothing in this approach, apart from the inevitable increased computational requirements, to prevent straightforward extensions to more complex control systems with more parameters.

5 Evolving the Model

Simulating the evolutionary process for our model involves taking a whole population of individual instantiations and allowing them to learn, procreate and die in a manner approximating these processes in real (biological) systems. The genotype of each new individual will depend only on the genotypes of its two parents and random mutation. Then during their life each individual will learn from their environment how best to adjust their weights to perform most effectively. Eventually, perhaps after producing a number of children, each individual dies. Obviously, in nature, or for complete physical robots, the ability of an individual to survive or reproduce will depend on a number of factors that are related in a complicated manner to that individual's performance on a range of related and unrelated tasks. For the purposes of our simplified model, however, I shall assume all other factors to be equal and consider it to be a sufficiently good approximation to take a simple linear relation between our single task fitness function and the survival or procreation fitness. In fact, any monotonic relation should result in similar evolutionary trends, but it is very easy to lose weak effects in the noise of the rather coarse simulations forced upon us by current computational resource limitations.

Given that, initially at least, we are aiming to replicate effects that arise in biological evolution, it is appropriate to follow a more natural approach to procreation, mutation and survival than has been used in many evolutionary simulations in the past (e.g. in Belew & Mitchell, 1996). If, as is often done, we were to train each member of the whole population for a fixed time and pick the fittest to breed and form the next generation, there would be no incentive for individuals to learn as quickly as possible, and efficient learners would not evolve. Instead, as in most biological systems, our populations contain competing learning individuals of all ages, each with the potential for dying or procreation at each stage. During each simulated year, every individual learns from their own experience with a new randomly generated common environment (i.e. set of training/testing data) and has their fitness measured. Random pairs of individuals are then forced to compete, with the least fit dying (i.e. being removed from the population). Additionally, a random subset of the oldest individuals die of old age. The dead are replaced by children, each having one parent who is the fittest of a randomly chosen pair from the remaining population, who randomly chooses their mate from the rest of whole population. Each child inherits characteristics from both parents such that each innate free parameter is chosen at random somewhere between the values of its parents, with sufficient noise (or mutation) that there is a reasonable possibility of the parameter falling outside the range spanned by the parents.

Note that even when a good set of innate weights have evolved, the control system will still benefit from being plastic since that will allow each individual to fine tune its performance after a noisy procreation process and/or being born into an unpredictable environment. Many biological systems also need plasticity to compensate for changes (e.g. growing size) that naturally take place during their own maturation period, and robots will need to compensate for plant drift (e.g. gear wear). For the current study, such changes were simulated by introducing a simple output scale factor that varies linearly from 0.5 to 1.0 over the first ten years of life for each individual. (It turns out that the precise details of this variation are not crucial.) In humans this maturation might correspond to changes in inter-pupilliary distance for the eye rotation system, or changes in arm length for reaching or pointing. The important consequence is that the appropriate innate/newborn weights will not be the same as the final adult values. The pattern of plasticities that evolve will allow the system to learn most efficiently how to optimize its weights throughout its life.

Ultimately, the simulations might benefit from more realistic encodings of the parameters, concepts such as recessive and dominant genes, learning and procreation

costs, different inheritance and mutation details, different survival and procreation criteria, more restrictive mate selection regimes, offspring protection, different learning algorithms and fitness functions, and so on, but for the purposes of this paper, our simplified approach proved adequate. An important consideration, however, was that limited computational resources will generally only permit rather coarse approximations of biological evolutionary processes, and it was important to fix the the various simulation parameters appropriately. For example, if all the individuals were able to learn how to perform the given task perfectly by the end of their first year, and we only tested their performance once per year, then the advantage of those that learn in two months over those that take ten is lost and the simulated evolution would not be very effective. Since the individuals were allowed to evolve their own learning rates, this had to be controlled by limiting the amount of training data each individual experienced in each year. Choosing a fixed population size of only 100 was a trade-off between maintaining genetic diversity and running the simulations reasonably quickly. The death rates were set in order to produce reasonable age distributions. This meant around 10 deaths per year due to competition, and another 4 individuals over the age of 30 dying each year due to old age. The procreation and mutation parameters were chosen to speed the evolution as much as possible without introducing too much noise into the process. Coding such a system in C typically resulted in around 1,000 simulated years per CPU hour on an average UNIX workstation. These details were kept constant across all the simulations I shall now present.

6   Simulation Results

Simulation results for a typical run of the basic system described above are shown in Figure 2. First we see that, although all the initial weights $\mathbf{W}(0)$ and learning rates $\mathbf{P}(0)$ were started with random values from the range $[0, 20]$, the population means quickly evolve to take on appropriate values. This demonstrates explicitly how genetic assimilation of learned behaviour (i.e. learned parameter values) will occur automatically, without Lamarckian inheritance, to reduce the inherent costs of learning (e.g. periods of poor performance). An important point to notice is the large variation in learning rates that emerge for the different components. Using a single learning rate for the whole system, as is common in neural network modelling, would clearly not be a good strategy in this type of system. The evolved parameters will result in good values for the weights throughout the individuals' lives. All the weights will need an initial fine tuning to remove the noise introduced by the procreation process, then some weights (WI and WP) need to adjust during the maturation period, while others (WT and WB) need little further change. The plots of $WP(t)$ and

$WT(t)$ for a typical evolved population show this quite clearly. We can see that even for individuals of the same age, there is still a degree of variability in the parameter values. The next graph shows this variability more clearly by plotting the standard deviations across the evolved population of the parameter values at each age, normalised by the corresponding adult values. These ranges of *individual differences* vary with age and ultimately depend on the sensitivity of the fitness function with respect to the corresponding parameters. This is discussed further in Bullinaria (2001a). Finally, we see how the plasticity scale factor $s(t)$ varies with age t. In particular, we see that the plasticity falls drastically between birth and the end of the maturation period, thus confirming that critical periods for learning will arise as a natural consequence of the evolutionary process.

The results from the basic system naturally lead to the question of what happens if an individual needs to be able to learn or adapt later in life, after the standard learning period is over. There is a traditional saying that "old dogs cannot learn new tricks", but it seems unlikely that evolution would allow the plasticities to decay away to small values in situations where late life adaptation is regularly required. To introduce such a requirement, the basic model was modified so that there was a sudden step in the output scale factor from 1.0 to 0.75 at the age of 20. (Again it turns out that the precise details of this variation are not crucial.) There is no need to specify whether this variation corresponds to an internal factor (e.g. compensation for system damage or deterioration) or an external factor (e.g. adaptation to changes in the operating environment), as they will have the same effect. Obviously, the need for real late life adaptation will rarely be so predictable, but the consequences for our model will be similar, and the simplification makes it easier to interpret the results.

Figure 3 shows how this changes the simulation results from those of the basic model in Figure 2. The most obvious difference is in the plot of $WP(t)$ where we see the required step change at age 20 has been learned successfully and quickly. We also see a corresponding blip in the individual differences due to variations in the adaptation process. The plot of $s(t)$ shows the initial fall as before, but then a peak to give the increased plasticity required at the age of 20. This gives us confidence that our evolutionary simulations really are picking up the requirement for plasticity, and not some confounding factor.

A final situation to consider, that often arises in biological development, is when one level of processing relies on signals from another sub-system. If the sub-system supplying those signals is not fully developed, it might be sensible to wait until it is before beginning to learn how to use its signals. For example, the adult human eye rotation (vergence) system uses an image disparity signal, and humans have to wait until 12-16 weeks of age
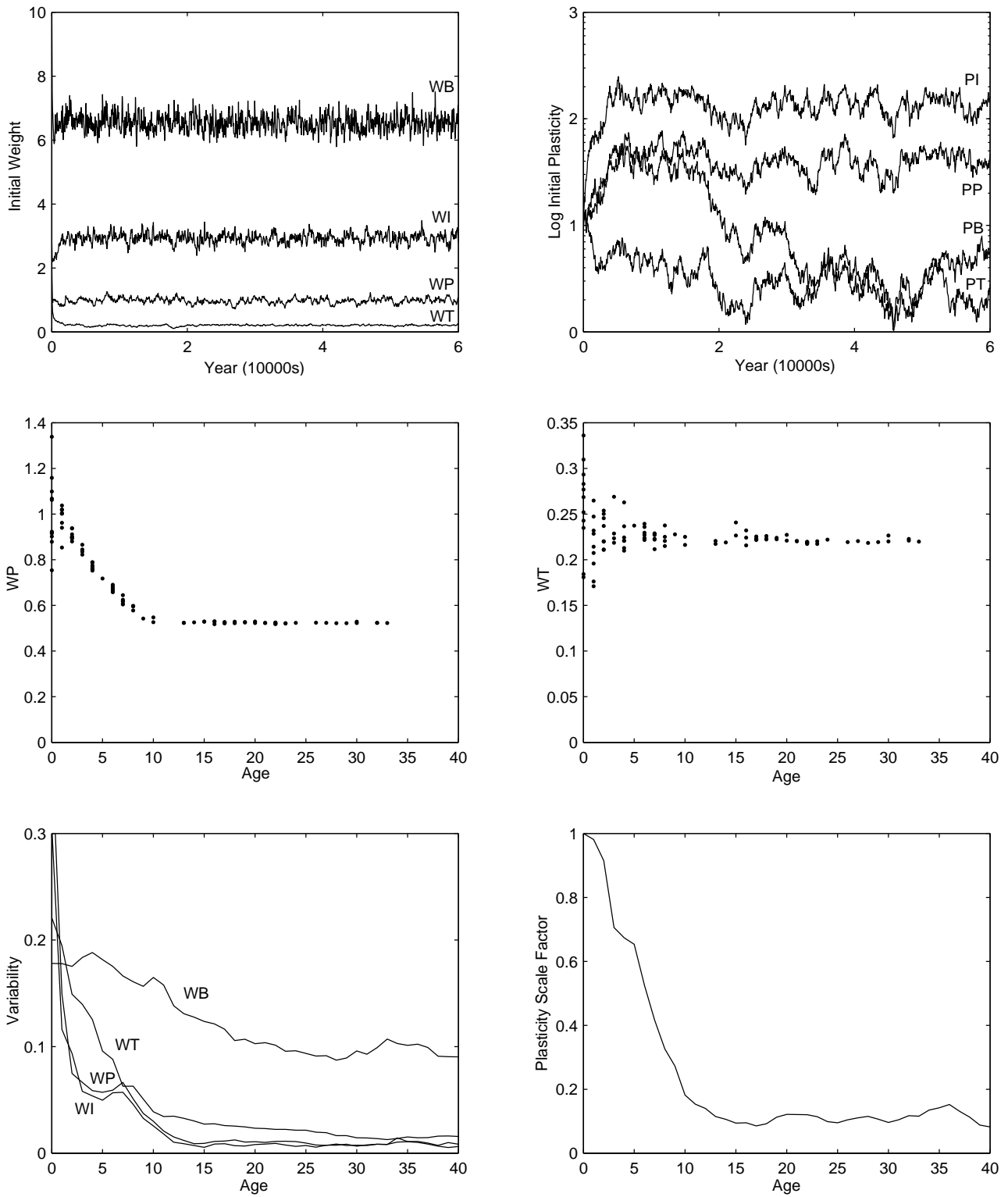
Figure 2: Evolution and learning in a typical simulation of the basic system. Appropriate initial weights and learning rates evolve quite quickly, and these result in good weights at each age. The remaining ranges of individual differences vary between the four weights depending on how crucial each is to the fitness, and on how quickly they are optimized by the learning. Individuals in the evolved population have plasticities that fall rapidly between birth and the end of their maturation period.
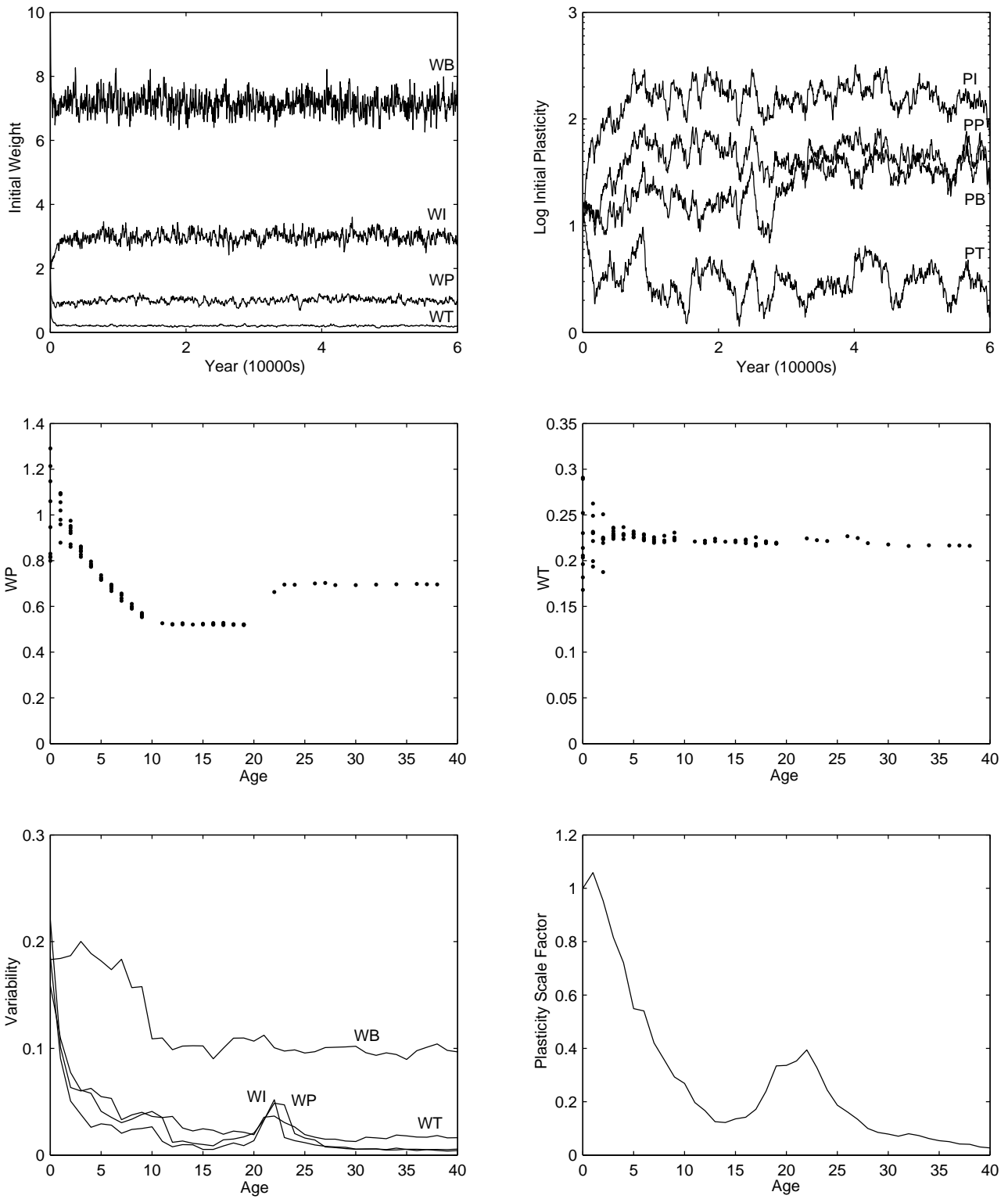
Figure 3: Evolution and learning in a typical simulation when late life adaptation is required. Again the initial weights and learning rates evolve quickly and appropriately to result in good weights at each age. Evolution here produces individuals in the final population with plasticities appropriate for the learning or adaptation that is forced upon them, with a peak around the age of 20 that enables them to cope efficiently with the changes required at that age.
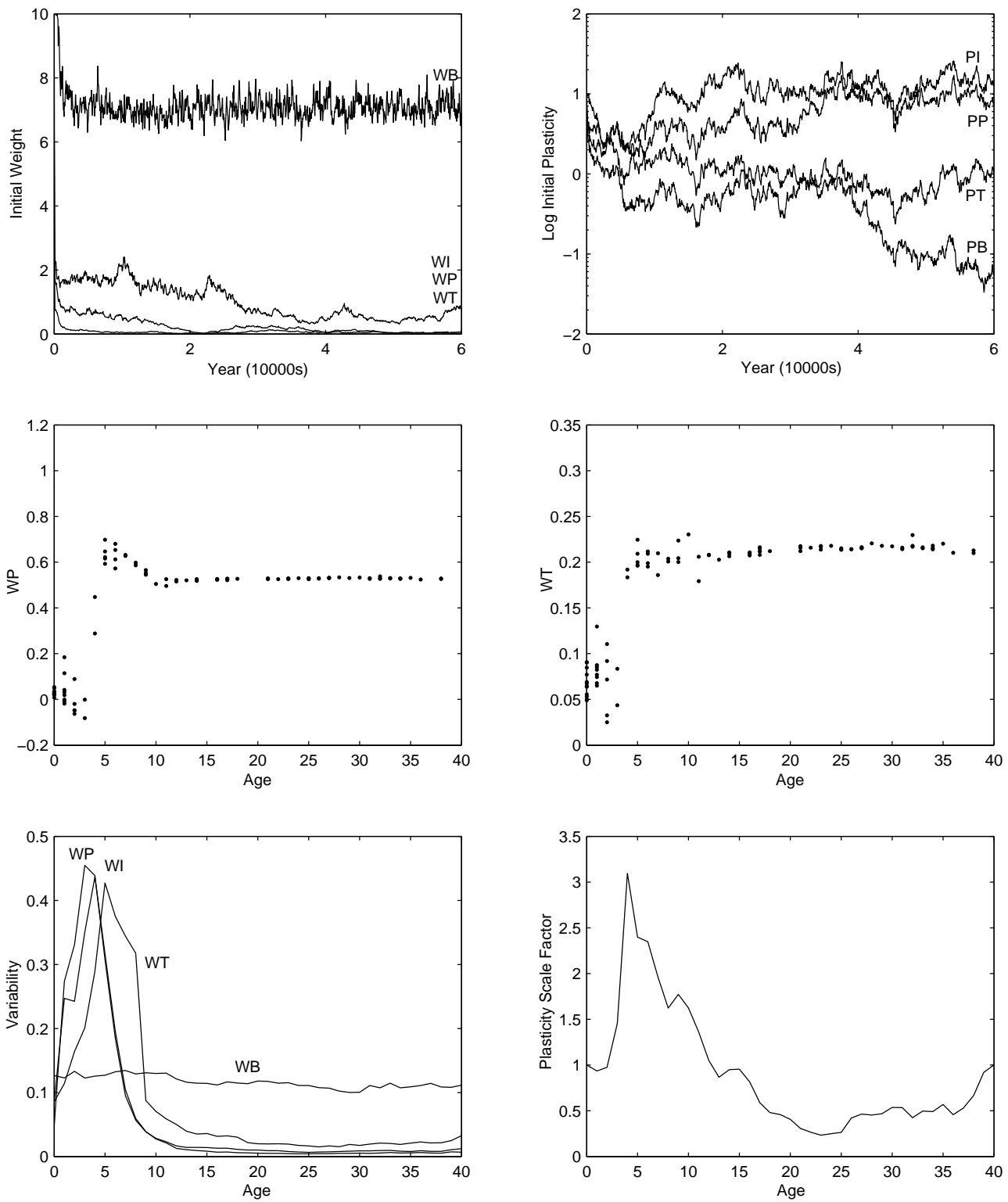
Figure 4: Evolution and learning in a typical simulation when there is a dependency on the development of lower level subsystems. Here, relatively low values emerge for the initial weights (apart from the bias WB) and initial learning rates, so that the system avoids processing or learning from signals that are just noise. This results in a population of individuals with critical periods for learning starting around age three, and a correspondingly different profile of individual differences.

before this signal relatively suddenly becomes available. To simulate such an effect in our basic model, the error signal was replaced by low level noise for each individual until they reached three years of age.

Figure 4 shows how this affects the standard results of Figure 2. The changes here are rather clear. First, the initial/innate weights WI, WP and WT all drop to very low values, leaving the system with an appropriate constant output driven by the bias WB, and no interference from the noisy input signal. Naturally, the initial learning rates are also all very low, because learning from noise is not a good strategy, but they quickly rise to coincide with the onset of the input signal at the age of three. By the age of nine, the system has caught up with the performance levels of Figure 2. Once again our biological style evolutionary approach leads to sensible patterns of initial parameters and plasticity variations.

There are two competing effects that determine what is an appropriate plasticity. In order to survive in competition with fitter adults and/or a hostile environment, a newborn needs to adapt as quickly as possible to its environment. It also needs to adapt efficiently to its own maturation. Large plasticities will be beneficial for both. In adults, however, large plasticities can lead to an unstable learning system, in which unusual/extreme experiences can potentially result in a large shift of the systems' parameters with a serious reduction in overall fitness. Lower learning rates in this situation will allow smoother optimal parameter estimation and more consistently good responses in a varied environment. The above simulations demonstrate explicitly how a process of evolution by natural selection can result in a population of individual systems that deal with these conflicting requirements by having plasticities that vary appropriately with age under normal maturation, when late life adaptation is required, and when there is a dependence on the prior development of other sub-systems.

7   Conclusions

We have seen how allowing populations of simple adaptable control systems to evolve by natural selection produces individuals that perform and adapt efficiently under a range of conditions. Along the way have been identified various advantages of following the processes of biological evolution more closely than is commonly the case. A Baldwin (1896) type interaction between learning and evolution results in the emergence of appropriate initial/innate values for each parameter/weight in the system, along with appropriate (different) learning rates for each parameter. We also observe a natural propensity for the evolution of leaning rates (plasticities) that vary sensibly with age, quite independently of any physical overheads of the plasticity. This is consistent with the well known "critical periods" of human brain development (Julesz & Kovacs, 1995). It is reasonable to expect that such an evolutionary approach will prove a profitable strategy for obtaining improved performance in systems for larger scale applications requiring adaptable controllers – such as robots.

References

Baldwin, J.M. (1896). A New Factor in Evolution. *The American Naturalist*, **30**, 441-451.

Belew, R.K. & Mitchell, M. (Eds) (1996). *Adaptive Individuals in Evolving Populations*. Reading, MA: Addison-Wesley.

Bullinaria, J.A. (2001a). Exploring the Baldwin Effect in Evolving Adaptable Control Systems. In: R.F. French & J.P. Sougne (Eds), *Connectionist Models of Learning, Development and Evolution*, 231-242. London: Springer.

Bullinaria, J.A. (2001b). Evolving Variable Plasticity in Neural Systems. In: N. Mastorakis (Ed), *Advances in Fuzzy Systems and Evolutionary Computation*, 236-241. World Scientific and Engineering Society Press.

Bullinaria, J.A. & Riddell, P.M. (2000). Learning and Evolution of Control Systems. *Neural Network World*, **10**, 535-544.

Bullinaria, J.A. & Riddell, P.M. (2001). Neural Network Control Systems that Learn to Perform Appropriately. *International Journal of Neural Systems*, **11**, 79-88.

Bullinaria, J.A., Riddell, P.M. & Rushton, S.K. (1999). Neural Network Modelling of the Linkages between Accommodation and Vergence. *Investigative Ophthalmology and Visual Science*, **40**, S362.

Greenough, W.T., Black, J.E. & Wallace, C.S. (1987). Experience and Brain Development. *Child Development*, **58**, 539-559.

Jacobs, R.A. (1988). Increased Rates Of Convergence Through Learning Rate Adaptation. *Neural Networks*, **1**, 295-307.

Julesz, B. & Kovacs, I. (Eds) (1995). *Maturational Windows and Adult Cortical Plasticity*. Reading, MA: Addison-Wesley.

Levine, W.S. (1996). *The Control Handbook*. London: CRC Press.

Nolfi, S. & Floreano, D. (2000). *Evolutionary Robotics*. Cambridge, MA: MIT Press.

Schor, C.M., Alexander, J., Cormack, L. & Stevenson, S. (1992). Negative Feedback Control Model of Proximal Convergence and Accommodation. *Ophthalmic and Physiological Optics*, **12**, 307-318.

Yao, X. (1999). Evolving Artificial Neural Networks. *Proceedings of the IEEE*, **87**, 1423-1447.