Proceedings of the British Academy, **88**, 77–93

Why Culture is Common, but Cultural Evolution is Rare

ROBERT BOYD* & PETER J. RICHERSON†
* *Department of Anthropology, University of California, Los Angeles, CA 90024, USA*
† *Division of Environmental Studies, University of California, Davis, CA 95616, USA*

Keywords: cultural evolution; social learning; dual inheritance; imitation.

Summary. If culture is defined as variation acquired and maintained by social learning, then culture is common in nature. However, cumulative cultural evolution resulting in behaviours that no individual could invent on their own is limited to humans, song birds, and perhaps chimpanzees. Circumstantial evidence suggests that cumulative cultural evolution requires the capacity for observational learning. Here, we analyse two models of the evolution of psychological capacities that allow cumulative cultural evolution. Both models suggest that the conditions which allow the evolution of such capacities when rare are much more stringent than the conditions which allow the maintenance of the capacities when common. This result follows from the fact that the assumed benefit of the capacities, cumulative cultural adaptation, cannot occur when the capacities are rare. These results suggest why such capacities may be rare in nature.

INTRODUCTION

CULTURAL VARIATION IS COMMON IN NATURE. In creatures as diverse as rats, pigeons, chimpanzees, and octopuses, behaviour is acquired through social learning. As a result, the presence of a particular behaviour in a population makes it more likely that individuals in the next generation will acquire the same behaviour which, in turn, results in persistent differences between populations that are not due to genetic or environmental differences.

In sharp contrast, cumulative cultural evolution is rare. Most culture in non-human animals involves behaviours that individuals can, and do, learn on their own. There are only a few well documented cases in which cultural change accumulates over many generations leading to the evolution of behaviours that no individual could invent—the only well documented examples are song dialects in birds, perhaps some behaviours in chimpanzees, and of course many aspects of human behaviour.

We believe that this situation presents an important evolutionary puzzle. The ability to accumulate socially learned behaviours over many generations has allowed humans to develop subtle, powerful technologies, and to assemble complex institutions that permit us to live in larger, and more complex societies than any other mammal species. These accumulated cultural traditions allow us to exploit a far wider range of habitats than any other animal, so that even with only hunting and gathering technology, humans became the most widespread mammal on earth. The fact that simple forms of cultural variation exist in a wide variety of organisms suggests that intelligence and social life alone are not sufficient to allow cumulative cultural evolution. Cumulative cultural change seems to require some special, derived, probably psychological, capacity. Thus we have the puzzle, if cultural traditions are such a potent means of adaptation, why is this capacity rare?

In this paper we suggest one possible answer to this question. We begin by reviewing the literature on animal social learning. We then analyse two models of the evolution of the psychological capacities that allow cumulative cultural evolution. The results of these models suggest a possible reason why such capacities are rare.

CULTURE IN OTHER ANIMALS

There has been much debate about whether other animals have culture. Some authors define culture in human terms. That is, the investigator essays human cultural behaviour and extracts a number of 'essential' features. For example Tomasello *et al.* (1993) argue that culture is learned by all group members, faithfully transmitted, and subject to cumulative change. Then to be cultural, the behaviour of other animals must exhibit these features. Moreover, a heavy burden of proof is placed on those who would claim culture for other animals—if there is any other plausible interpretation, it is preferable. Others (McGrew 1992; Boesch 1993) argue that a double standard is being applied. If the behavioural variation observed among chimpanzee populations were instead observed among human populations, they argue, anthropologists would regard it as cultural.

Such debates make little sense from an evolutionary perspective. The psychological capacities that underpin human culture must have homologies in the brains of other primates, and perhaps other mammals as well. Moreover, the functional significance of social transmission in humans could well be related to its functional significance in other species. The study of the evolution of human culture must be based on categories that allow human cultural behaviour to be compared to potentially homologous, functionally related behaviour of other organisms. At the same time, such categories should be able to distinguish between human behaviour and the behaviour of other organisms because it is quite plausible that human culture is different in important ways from related behaviour in other species.

Here we define cultural variation as differences among individuals that exist because they have acquired different behaviour as a result of some form of social learning. Cultural variation is contrasted with genetic variation, differences between individuals that exist because they have inherited different genes from their parents, and environmental variation, differences between individuals due to the fact that they have experienced different environments. Cultural variation is often lumped together with environmental variation. However, as we have argued at length elsewhere (Boyd & Richerson 1985), this is an error. Because cultural variation is transmitted from individual to individual it is subject to population dynamic processes analogous to those that effect genetic variation and quite unlike the processes that govern other environmental effects. Combining cultural and environmental effects into a single category conceals these important differences.

There is much evidence that cultural variation, defined this way, is very common in nature. In a review of social transmission of foraging behaviour, Levebre & Palameta (1988) give 97 examples of cultural variation in foraging behaviour in animals as diverse as baboons, sparrows, lizards, and fish. Song dialects are socially transmitted in many species of songbirds. Three decades of study shows that chimpanzees have cultural variation in subsistence techniques, tool use, and social behaviour (Wrangham *et al.* 1994; McGrew 1992).

There is little evidence, however, of cumulative cultural evolution in other species. With a few exceptions, social learning leads to the spread of behaviours that individuals could have learned on their own. For example, food preferences are socially transmitted in rats. Young rats acquire a preference for a food when they smell the food on the pelage of other rats (Galef 1988). This process can cause the preference for a new food to spread within a population. It can also lead to behavioural differences among populations living in the same environment, because current foraging behaviour depends on a history of social learning. However, it does not lead to the cumulative evolution of new, complex behaviours that no individual rat could learn on its own.

In contrast, human cultures do accumulate changes over many generations, resulting in culturally transmitted behaviours that no single human individual could invent on their own. Even in the simplest hunting and gathering societies people depend on such complex, evolved knowledge and technology. To live in the arid Kalahari, the !Kung San need to know what plants are edible, how to find them during different seasons, how to find water, how to track and find game, how to make bows and arrow poison, and many other skills. The fact that the !Kung can acquire the knowledge, tools, and skills necessary to survive the rigors of the Kalahari is not so surprising—many other species can do the same. What is amazing is that the same brain that allows the !Kung to survive in the Kalahari, also permits the Inuit to acquire the very different knowledge, tools, and skills necessary to live on the tundra and ice north of the Arctic circle, and the Ache the knowledge, tools, and skills necessary to live in the tropical forests of Paraguay. There is no other animal that occupies a comparable range of habitats or utilizes a comparable range of subsistence techniques and social structures. Two kinds of evidence indicate that such differences result from cumulative cultural evolution of complex traditions. First, such gradual change is documented in both the historical and archaeological records. Second, cumulative change leads to a branching pattern of descent with modification in which more closely related populations share more derived characters than distantly related populations. Although the possibility of horizontal transmission among cultural lineages makes reconstructing such cultural phylogenies difficult for ‘cultures’ (Boyd *et al.* in press), patterns of cultural descent can be reconstructed for particular cultural components, such as language or technologies.

Circumstantial evidence suggests that the ability to acquire novel behaviours by observation is an essential for cumulative cultural change. Students of animal social learning distinguish *observational learning or true imitation*, which occurs when younger animals observe the behaviour of older animals and learn how to perform a novel behaviour by watching them, from a number of other mechanisms of social transmission which also lead to behavioural continuity without observational learning (Galef 1988; Visalberghi & Fragazy, 1990; Whiten & Ham 1992). One such mechanism, *local enhancement*, occurs when the activity of older animals increases the chance that younger animals will learn the behaviour on their own. If younger, naive individuals are attracted to the locations in the environment where older, experienced individuals are active they will tend to learn the same behaviours as the older individuals. Young individuals do not acquire the information necessary to perform the behaviour by observing older individuals. Instead, the activity of others causes them to be more likely to acquire this information through interaction with the environment. Imagine a young monkey acquiring its food preferences as it follows its mother around. Even if the young monkey never pays any attention to what its mother eats, she will lead it to locations where some foods are common and others rare, and the young monkey may learn to eat much the same foods as mom.

Local enhancement and observational learning are similar in that they both can lead to persistent behavioural differences among populations, but only observational learning allows cumulative cultural change (Tomasello *et al.* 1993). To see why, consider the cultural transmission of stone tool use. Suppose that on their own in especially favourable circumstances, an occasional early hominid learned to strike rocks together to make useful flakes. Their companions, who spent time near them, would be exposed to the same kinds of conditions and some of them might learn to make flakes too, entirely on their own. This behaviour could be preserved by local enhancement because groups in which tools were used would spend more time in proximity to the appropriate stones. However, that would be as far as it would go. Even if an especially talented individual found a way to improve the flakes, this innovation would not spread to other members of the group because each individual learned the behaviour anew. Local enhancement is limited by the learning capabilities of individuals, and the fact that each new learner must start from scratch. With observational learning, on the other hand, innovations can persist as long as younger individuals are able to acquire the modified behaviour by observational learning. To the extent that observers can use the behaviour of models as a starting point, observational learning can lead to the cumulative evolution of behaviours that no single individual could invent on its own.

Most students of animal social learning believe that observational learning is limited to humans, and perhaps, chimpanzees and some bird species. Several lines of evidence suggest that observational learning is not responsible for cultural traditions in other animals. First, many of the behaviours, like potato washing in Japanese macaques, are relatively simple, and could be learned independently by individuals in each generation. Second, new behaviours like potato washing often take a long time to spread through the group, a pace more consistent with the idea that each individual had to learn the behaviour on its own. Finally, extensive laboratory experiments capable of distinguishing observational learning from other forms of social transmission like local enhancement have usually failed to demonstrated observational learning (Galef 1988; Whiten & Ham 1992; Tomasello *et al.* 1993; Visalberghi 1993), except in humans and song birds. (In many song birds, song traditions are transmitted by imitation, but little or nothing else is.) The fact that observational learning appears limited to humans seems to confirm that observational learning is necessary for cumulative cultural change. However, one must be cautious here because most students of animal social learning refuse to invoke observational learning unless all other possible explanations have been excluded. Thus, there actually may be many cases of observational learning that are interpreted as social enhancement or some putatively simpler mechanism. A few well controlled laboratory studies do apparently show some true imitation in non-human animals (Heyes 1993; Dawson & Foss 1965), and striking anecdotes suggest that observational learning may occur in organisms as diverse as parrots (Pepperberg 1988) and orangutans (Russon & Galdikas 1993).

Adaptation by cumulative cultural evolution is apparently not a by-product of intelligence and social life. Cebus monkeys are among the world's cleverest creatures. In nature, they use tools and perform many complex behaviours, and in captivity, they can be taught extremely demanding tasks. Cebus monkeys live in social groups and have ample opportunity to observe the behaviour of other individuals of their own species. Yet good laboratory evidence suggests that cebus monkeys make no use of observational learning. This suggests that observational learning is not simply a by-product of intelligence and the opportunity to observe conspecifics. Rather, observational learning seems to require special psychological mechanisms (Bandura 1986). This conclusion suggests, in turn, that the psychological mechanisms that enable humans to learn by observation are adaptations have been shaped by natural selection because culture is beneficial. Of course, this need not be the case. Observational learning could be a by-product of some other adaptation that is unique to humans, such as bipedalism, dependence on complex vocal communication, or the capacity for deception. However, given the great importance of culture in human affairs, it is reasonable to think about the possible adaptive advantages of culture. In what follows we consider the two mathematical models of the evolution of the capacity for observational learning based on this assumption.

MODELS OF THE EVOLUTION OF SOCIAL LEARNING

The maintenance of cultural variation involves two quite different processes (Figure 1). First, there must be some kind of *transmission* of information

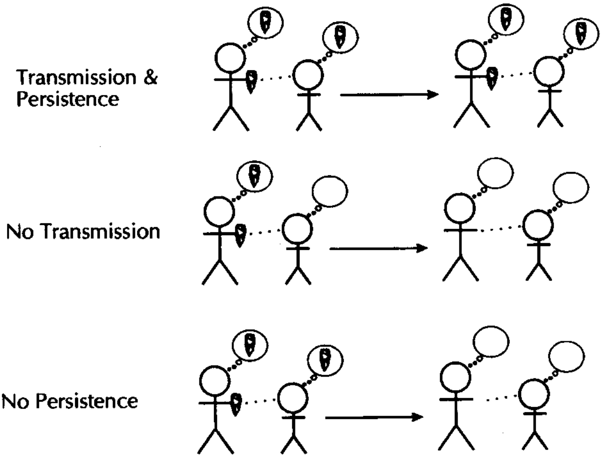

Figure 1. The maintenance of cultural transmission requires both the accurate transmission of mental representations from experienced to inexperienced individuals, and the persistence of those representations through the lives of individuals until such time that they act as models for others.

from one brain to another. Consider, for example, the maintenance of the use of a particular kind of tool. Individuals have information stored in their brain that allows them to manufacture and use the tool. For use of the tool to persist through time, observing tool use and manufacture must cause individuals in the next 'generation' to acquire information that allows them to manufacture and use the same tool. (We put generation in quotes because the same model can be used to represent culture change occuring on much shorter time scales. See Boyd & Richerson 1985: 68–69.) As we have seen, this transmission may occur because individuals can learn how to make and use tools by observation, or because observation stimulates them to learn on their own how to make and use the tool, for example by local enhancement. Second, individuals must preserve the information that allows them to make and use the tool until such time that they serve as models for the next generation of individuals. Such *persistence* may fail to occur for two different reasons: individuals may forget how to make or use the tool, or they may, as a result of interacting with the environment, modify the information stored in their brains so that they make or use the tool in a significantly different way. Without both transmission and persistence there cannot be culturally transmitted variation.

Our previous work on the evolution of culture (Boyd & Richerson 1985, 1988, 1989, in press) has focused on the evolution of persistence. All of the models analysed in these studies assume that transmission occurs, and consider the evolution of genes that affect the extent to which behaviour acquired by imitation is modified by individual learning. They differ in how the trait is modelled (discrete vs. continuous), how environmental variation is modelled, whether individuals are sensitive to the number of models who exhibit a particular cultural variant, and a number of other features. This work leads to the robust conclusion that natural selection will favour individuals who do not modify culturally acquired behaviour when individual learning is costly or error prone, and environments are variable, but not too variable. Thus, natural selection can favour persistence. (See Rogers 1989 for a related model.)

In several papers, Feldman and his co-workers (Cavalli-Sforza & Feldman 1983a, 1983b; Aoki & Feldman 1987) have considered the evolution of genes that affect transmission. In these models it is assumed that there is a beneficial trait that can only be acquired by cultural transmission, not by individual learning. They further allow for the possibility that successful transmission requires new behaviour both on the part of the individual acquiring the behaviour and in the individual modelling the behaviour. Thus there are two different genetic loci, one affecting the behaviour of the transmitter and a second affecting the behaviour of the receiver. For transmission to evolve, there must be substitutions at both loci. These models are very relevant to the evolution of communication systems. However, they cannot address the questions posed here because the culturally transmitted trait cannot be acquired or modified by individual learning.

Here we consider two models of the evolution of psychological capacities that allow the transmission of behaviour that can be acquired or modified through individual learning. Each model is designed to answer the same basic question: What are the conditions under which selection can favour a costly psychological capacity that allows individuals to acquire behaviour by imitation? The primary difference between the models is the nature of the culturally transmitted behaviour. In the first model, the behaviour is discrete—individuals are either skilled or unskilled, and the skill can be acquired either by social or individual learning. In the second model, there is a continuum of behaviours subject to stabilizing selection. Only the continuous trait model allows true cumulative cultural change leading to behaviours that individuals cannot learn on their own. However, the discrete model allows us to investigate the effects of several factors that are difficult to include in the continuous character model. As we will see, both models tell a similar story about why there is a selective barrier to the evolution of the capacity for observational learning, and why capacities that allow local enhancement and related mechanisms do not face a similar barrier.

Discrete character model

Consider an organism that lives in a temporally variable environment that can be in an infinite number of states. In each state, individuals can acquire a skill which increases fitness, so that unskilled individuals have fitness W_0, and skilled individuals have fitness $W_0 + D$. Each generation there is a probability γ that the environment switches from its current state to a different state. When this occurs, the old skill is no longer useful in the new environment.

There are two genotypes with different learning rules. *Individual learners* acquire the skill appropriate to the current environment with probability δ at a cost C_I. *Social learners* observe n randomly selected members of the previous generation. If there is a skilled individual among the n, an imitator acquires the skill at cost C_S. Otherwise they acquire the skill with probability δ at a cost C_I. The ability to acquire the skill by social learning reduces the fitness of an individual an amount K. Thus, parameters C_I and C_S give the variable costs of individual and social learning respectively, and K gives the fixed cost associated with the capacity for social learning.

It is shown in the appendix that social learning can increase when rare, and is the only ESS when the following condition holds

$$(1-(1-\delta)^n)(1-\gamma)[D(1-\delta)+C_I-C_S] > K \tag{1}$$

When (1) is true, social learning has higher fitness than individual learning no matter what the mix of the two types in the population. The term in square brackets gives the fitness benefit of acquiring the skill through social rather than individual learning—$C_I - C_S$ is the advantage that results from the fact that social learning may reduce the cost of acquiring the trait, and $D(1-\delta)$ is the advantage that results from being more likely to acquire the skill. Sensibly, the latter term implies that the fitness advantage of social learning increases as the likelihood that individuals will learn the trait on their own, δ, decreases. The less likely it is that individual learners will acquire the skill, the bigger the relative advantage that accrues to social learning. The fitness benefit is discounted by the two factors on the left hand side of expression (1). The term $1-\gamma$ expresses the fact that social learning is only beneficial if the environment has not changed, and term $1-(1-\delta)^n$ gives the probability that at least one of the n individuals from the previous generation will have acquired the behaviour when social learning is rare. Notice that this latter term decreases as the probability of learning the trait decreases. Thus the net advantage of social learning is highest at intermediate values of δ, when there is a good chance that individuals will learn the skill on their own, but also a good chance that they won't.

When (1) is not satisfied, there is a range of conditions in which social learning cannot increase when rare, but is an ESS once it becomes common. In this analysis we are limited to the case $n = 1$ because when $n > 1$ the dynamics of the cultural traits are nonlinear, and such systems are difficult to analyse in autocorrelated random environments. With this assumption, social learning is an ESS when:

$$\frac{\delta(1-\gamma)(D(1-\delta)+C_I-C_S)}{\gamma+(1-\gamma)\delta} > K \tag{2}$$

To compare this expression with (1), notice that when $n = 1$, $1-(1-\delta)^n = \delta$, and thus, the benefit of social learning when it is common is the benefit when rare divided by the term $\gamma+(1-\gamma)\delta$. When individual learners are likely to acquire the skill (so that δ is large), the conditions for social learning to increase when rare (1) and to persist when common (2) will be similar. However, when individual learners are unlikely to acquire the skill ($\delta \ll 1$) and the rate of environmental change is slow ($\gamma \ll 1$), social learning will be able to persist when common under a much wider range of conditions than it can increase when it is rare. When social learning is rare, most of the population will be individual learners who have little chance of acquiring the skill. As a consequence, social learning will provide little benefit because there will be few skilled individuals to observe. When social learning is common, the population will slowly accumulate the skill over many generations. If the environment does not change too often, the social learning population will spend most of the time with the skill at high frequency, and thus the cost of the capacity for social learning need only be less than the net benefit of acquiring the skill by individual learning.

Continuous character model

Consider an organism that is characterized by a single quantitative character that is subject to stabilizing selection. During generation t the optimum value of the quantitative character is θ_t. Each generation there is a probability γ that the environment changes. If the environment does not change then $\theta_{t+1} = \theta_t$. If it does change, then θ_{t+1} is a normal random variable with mean Θ, and variance H. Notice that this assumption implies that Θ is the long run optimum trait value.

Each individual acquires its trait value through a combination of genetic transmission, imitation, and individual learning. The adult trait value, x, is given by:

$$x = (1-a)[(1-i)\Theta + iy] + a\theta_t \tag{3}$$

The term $(1 - i)\Theta + iy$ represents a 'norm of reaction' which forms the basis for subsequent individual learning. It is acquired as the result of a combination of a genetically acquired norm of reaction at the long run optimum, Θ, and the observed trait value, y, of a randomly selected member of the previous generation. The parameter i governs the relative importance of genetic inheritance and imitation in determining the norm of reaction. When $i = 0$, the norm of reaction is completely determined by an innate, genetically inherited value. As i increases, the observed trait value of another individual has greater influence on the trait until, when $i = 1$, the norm of reaction is completely determined by observational learning. Because observational learning is assumed to require special purpose cognitive machinery, individuals incur a fitness cost proportional to the importance of observational learning in determining their norm of reaction, iC. Thus, C measures the incremental cost of the capacity for observational learning. Individuals adjust their adult behaviour from the norm of reaction toward the current optimum a fraction a. To capture the idea that cumulative change is possible we assume that a is small, so that the repeated action of learning and social transmission can lead to fitness increases that could not be attained by individual learning.

With these assumptions it is shown in the appendix that a population in which most individuals do not imitate can be invaded by rare individuals who imitate a little bit only if

$$(1 - \gamma)aH > C \tag{4}$$

The parameter H is a measure of how far the population is from the optimum in fitness units, on average, immediately after an environmental change. Since a population without imitation always starts from the same norm of reaction, Θ, the term aH is a measure of the average fitness improvement due to individual learning in a single generation. Thus, (4) says that imitation can evolve only when the benefit of imitating what individuals can learn on their own is sufficient to compensate for the costs of the capacity to imitate.

In contrast, the condition for social learning to be maintained once it is common is much more easily satisfied. It is shown in the appendix that a population in which $i = 1$ can resist invasion by rare alleles that reduce the reliance on imitation whenever:

$$\frac{(1 - \gamma)aH}{\gamma + (1 - \gamma)a} > C \tag{5}$$

If the rate at which the population adapts by individual learning, a, is greater than the rate at which the environment changes, γ, then a population in which social learning is common spends most of its time with the mean behaviour near the optimum. Thus, (5) says that imitation is evolutionarily stable as long as the cost of the capacity is less than a substantial fraction of the total improvement in fitness due to many generations of social learning.

DISCUSSION

Both of these models tell a similar story about the evolution of capacities that allow social learning. When social learning is rare, the only useful behaviour that is present in the population, and thus the only behaviour that can be acquired by social learning, is behaviour that individuals can learn on their own. In contrast, when social learning is common the population accumulates adaptive behaviour over many generations, and, as long as the environment does not change faster than adaptive behaviour accumulates, social learning allows individuals to acquire behaviours that are much *more* adaptive than they could acquire on their own.

This result provides a potential explanation for why cultural variation is so common in nature, but cumulative cultural evolution so rare. Capacities that increase the chance that individuals will learn behaviours that they could learn on their own will be favoured as long as they are relatively cheap. On the other hand, even though the benefits of cumulative cultural evolution are potentially substantial, selection cannot favour a capacity for observational learning when rare. Thus unless observational learning substantially reduces the cost of individual learning, it will not increase because there is an 'adaptive valley' that must be crossed before benefits of cumulative cultural change are realized. This argument suggests, in turn, that it is likely that the capacities that allow the initial evolution of observational learning must evolve as a side effect of some other adaptive change. For example, it has been argued that observational learning requires that individuals have what psychologists and philosophers call a 'theory of mind (Cheney & Seyfarth 1990; Tomasello *et al.* 1993).' That is, imitators must be able to understand that others have different beliefs and goals than they. Lacking such a theory, typical animals cannot make a connection between the acts of other animals and their own goal states, and thus can't interpret the acts of other animals as acts they might usefully perform. A theory of mind may have initially evolved to allow individuals to better predict the behaviour of other members of their social group. Once it had evolved for that reason it could be elaborated because it allowed observational learning and cumulative cultural evolution.

REFERENCES

Aoki, K. & Feldman, M.W. 1987: Toward a theory for the evolution of cultural communication: Coevolution of signal transmission and reception *Proceedings of the National Academy of Sciences, U.S.A.* 84, 7164–8.

Bandura, A. 1986: *Social Foundations of Thought and Action: A Social Cognitive Theory.* Englewood Cliffs, NJ: Prentice Hall.

Boesch, C. 1993: Aspects of transmission of tool-use in wild chimpanzees. In *Tools, Language, and Cognition in Human Evolution* (ed. K. R. Gibson & T. Ingold) pp. 171–189. Cambridge University Press, Cambridge

Boyd, R. & Richerson, P.J. 1985: *Culture and the Evolutionary Process.* University of Chicago Press, Chicago.

Boyd, R. & Richerson, P.J. 1988: An evolutionary model of social learning: the effects of spatial and temporal variation. In *Social Learning, Psychological and Biological Perspectives* (ed. T. Zentall & B.G. Galef), pp. 29–48. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ.

Boyd, R. & Richerson, P.J. 1989: Social Learning as an Adaptation. *Lectures on Mathematics in the Life Sciences* 20, 1–26.

Boyd, R. & Richerson, P.J. 1995: Why does culture increase human adaptability? *Ethology and Sociobiology* 16, 125–141.

Boyd, R., Richerson, P.J., Borgerhoff Mulder, M. & Durham, W.H. (In press) Are cultural phylogenics possible? In *Human Nature, Between Biology and the Social Sciences* (ed. P. Weingart, P.J. Richerson, S. Mitchell & S. Maasen).

Cavalli-Sforza, L.L. & Feldman, M.W. 1983a: Paradox of the evolution of communication and of social interactivity. *Proceedings of the National Academy of Sciences, U.S.A.* 80, 2017–2021.

Cavalli-Sforza, L.L. & Feldman, M.W. 1983b: Cultural versus genetic adaptation. *Proceedings of the National Academy of Sciences, U.S.A.* 80, 4993–4996.

Cheney, D. & Seyfarth, R. 1990: *How Monkeys See the World.* University of Chicago Press, Chicago.

Dawson, B.V. & Foss, B.M. 1965: Observational learning in budgerigars. *Animal Behaviour* 13, 470–474.

Galef, B.G. 1988: Imitation in animals: History, definitions, and interpretation of data from the psychological laboratory. In *Social Learning, Psychological and Biological Perspectives* (ed. T. Zentall & B.G. Galef, Jr.), pp. 3–29. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey.

Heyes, C.M. 1993: Imitation, culture and cognition. *Animal Behaviour* 46, 999–1010.

Levebre, L. & Palameta, B. 1988: Mechanisms, ecology, and population diffusion of socially-learned, food-finding behaviour in feral pigeons. In *Social Learning, Psychological and Biological Perspectives* (ed. T. Zentall & B.G. Galef, Jr.), pp. 141–165. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey.

McGrew, W. 1992: *Chimpanzee Material Culture.* Cambridge University Press, Cambridge.

Pepperberg, I. 1988: The importance of social interaction and observation in the acquisition of communicative competence: Possible parallels between avian and human learning. In *Social Learning, Psychological and Biological Perspectives* (ed. T. Zentall & B.G. Galef, Jr.), pp. 3–29. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey.

Rogers, A.R. 1989: Does biology constrain culture? *American Anthropologist* 90, 819–831.

Russon, A.E. & Galdikas, B. 1993: *Journal of Comparative Psychology.*

Tomasello, M., Kruger, A.C. & Ratner, H.H. 1993: Cultural learning. *Behavioral and Brain Sciences* 16, 495–552.

Visaberghi, E. 1993: Capuchin monkeys: A window into tool use in apes and humans. In *Tools, Language, and Cognition in Human Evolution* (ed. K.R. Gibson & T. Ingold), pp. 138–150. Cambridge University Press, Cambridge.

Visalberghi, E. & Fragazy, D.M. 1990: Do monkeys ape? In *Language and Intelligence in Monkeys and Apes* (ed. S. Parker & K. Gibson), pp. 247–273. Cambridge University Press, Cambridge.

Whiten, A. & Ham, R. 1992: On the nature and evolution of imitation in the animal kingdom: A reappraisal of a century of research. In *Advances in the Study of Behavior, Vol. 21* (ed. P.J.B. Slater, J.S. Rosenblatt, C. Beer & M. Milkinski), pp. 239–283. Academic Press, New York.

Wrangham, R.W., McGrew, W.C., DeWaal, F.B.M. & Heltne, P.G. 1994: *Chimpanzee Cultures*. Havard University Press, Cambridge.

APPENDIX

Analysis of discrete character model

Individual learners always have the same fitness:

$$W_l = W_0 + \delta D - C_l. \tag{A1.1}$$

The expected fitness of social learners depends on the frequency of social learners in the previous generation, q, the frequency of skilled individuals among social learners, p, and whether the environment has changed during the previous generation.

$$\begin{aligned} W_S = {} & \gamma(W_0 + \delta D - C_l) \\ & +(1-\gamma)(W_0 + \pi(D - C_s) + (1-\pi)(\delta D - C_l)) \end{aligned} \tag{A1.2}$$

where π is the probability that at least one of the n individuals in the sample of models has acquired the skill favoured in the previous environment, and can be calculated as below:

$$\pi = \sum_{i=0}^{n} \binom{n}{i} q^i (1-q)^{n-i} [1 - (1-p)^i (1-\delta)^{n-i}]. \tag{A1.3}$$

To understand this expression assume that there are i social learners among the n models observed by a given, naive social learner. The probability that all i of the social learners are not skilled is $(1-p)^i$, and the probability that the remaining $n-i$ individual learners are not skilled is $(1-\delta)^{n-i}$, and therefore, the probability that there is at least one skilled individual among the n given that there are i social learners is $1-(1-p)^i(1-\delta)^{n-i}$. Then to calculate π take the expectation over all values of i.

Thus, social learners will have higher fitness in a particular generation if

$$W_S - W_l = \pi(1-\gamma)(D(1-\delta) + C_l - C_S) - K > 0. \tag{A1.4}$$

We consider two special cases. Case 1: $q \approx 0$, $\pi \approx 1-(1-\delta)^n$. When social learners are rare, they will observe only individual learners, and thus the probability of observing at least one skilled individual does not depend on q or p. Thus, social learning will increase when rare if

$$(1-(1-\delta)^n)(1-\gamma)(D(1-\delta)+C_l-C_S)-K>0. \tag{A1.5}$$

Immediately after an environmental change, the frequency of skilled individuals among social learners is δ, and then increases monotonically until the next environmental change. Thus the expected value of π is greater than $(1-(1-\delta)^n)$, and if social learning can increase when rare it will continue to increase until it reaches fixation.

Case 2: $n=1$, $\pi = 1-q(1-p)-(1-q)(1-\delta)$. Assume that selection is sufficiently weak so that the effect of selection on cultural evolution can be ignored (i.e., on dynamics of p), and genetic evolution (the dynamics of q) responds to the stationary distribution of p.

Then the frequency of the currently favoured behaviour after learning and imitation is

$$p' = \begin{cases} \delta & \text{if environment changes} \\ (qp+(1-q)\delta)(1-\delta)+\delta & \text{if environment does not change.} \end{cases} \tag{A1.6}$$

Suppose at some time t the probability density for p is $f_t(p)$ with mean P_t. Then the mean of $f_{t+1}(p)$ given by

$$P_{t+1} = \int [(1-\gamma)((qp+(1-q)\delta)(l-\delta)+\delta)+\gamma\delta] f_t(p)dp. \tag{A1.7}$$

Integrating yields the following recursion for P_t

$$P_{t+1} = \gamma\delta + (1-\gamma)[(qP_t+(1-q)\delta)(1-\delta)+\delta]. \tag{A1.8}$$

Thus the equilibrium value of mean frequency of the favoured behaviour is:

$$P = \frac{\delta+(1-\gamma)(1-q)\delta(1-\delta)}{1-(1-\gamma)(1-\delta)q}. \tag{A1.9}$$

Assume that selection is weak enough that the dynamics of q respond to the stationary distribution of p. Then, since the expression for W_s is linear in p when $n=1$, we can substitute P for p. With this assumption

$$\pi = \frac{\delta}{1-(1-\gamma)(1-\delta)q}. \tag{A1.10}$$

Notice that $\pi > \delta$, which implies that social learners are more likely on average to acquire the skill. Substituting A1.10 into A1.4 yields the following condition for social learning to increase in frequency

$$\frac{(1-\gamma)(D(1-\delta)+C_l-C_S)\delta}{1-(1-\gamma)(1-\delta)q} > K. \tag{A1.11}$$

Analysis of continuous character model

Since we are free to determine the scale of measurement of trait values, we can, without loss of generality set $\Theta = 0$. Then the mean value of x in the population during generation t, X_t, is:

$$X_t = (1 - a)iX_{t-1} + a\theta_t. \tag{A2.1}$$

The logarithm of the fitness of an individual with adult trait value x is proportional to:

$$\ln(W) \propto -(x - \theta_t)^2 - C(i). \tag{A2.2}$$

Thus the expected fitness of an individual whose behavioural acquisition is governed by the parameter i is

$$E\{\ln(W)\} \propto -(1 - a)^2 E\{(iX_{t-1} - \theta_t)^2\} - C(i). \tag{A2.3}$$

Consider the competition between two genotypes. The common type has development characterized by parameter i and the rare type by $i + \delta$, where δ is very small. If one assumes that changes in i have no effect on the variance of the trait among the invading type individuals, the expected fitness of the invading type is approximately proportional to

$$E\{\ln(W)\} \propto -(1 - a)^2$$

$$\times[(i^2 + 2i\delta)E\{X_{t-1}^2\} - 2(i + \delta)E\{X_{t-1}\theta_t\} + \theta_t^2] - C(i) - \frac{\partial C}{\partial i}\delta. \tag{A2.4}$$

Combining expression A2.3 and A2.4 shows that the invading type will increase in frequency if

$$-(1 - a)^2[2i\delta E\{X_{t-1}^2\} - 2\delta E\{X_{t-1}\theta_t\}] - \frac{\partial C}{\partial i}\delta > 0. \tag{A2.5}$$

To calculate $E\{X_{t-1}\theta_t\}$ first notice that

$$\theta_t = \begin{cases} \theta_{t-1} & \text{with probability } 1-\gamma \\ \varepsilon & \text{with probability } \gamma \end{cases} \tag{A2.6}$$

where ε is an independent normal random variable with mean zero and variance H. Thus it follows that

$$E\{\theta_t X_{t-1}\} = (1 - \gamma)E\{\theta_{t-l}X_{t-1}\} + \gamma E\{X_{t-1}\varepsilon\}. \tag{A2.7}$$

Multiply both sides of A2.1 by θ_t and taking the expectation with respect to the joint stationary distributions yields:

$$E\{\theta_t X_t\} = (1 - a)iE\{\theta_t X_{t-1}\} + aH. \tag{A2.8}$$

Combing A2.8 and A2.9 yields the following expression for $E\{X_{t-1}\theta_t\}$:

$$E\{X_{t-1}\theta_t\} = \frac{(1-\gamma)aH}{1-i(1-\gamma)(1-a)}. \tag{A2.9}$$

To calculate $E\{X_{t-1}^2\}$ square both sides of A2.1, take the expectation, and using A2.9 solve:

$$E\{X_{t-1}^2\} = \frac{a^2 - 2i(1-a)E\{X_{t-1}\theta_t\}}{1-i^2(1-a)^2}. \tag{A2.10}$$

Substituting A2.9 and A2.10 into A2.5 and simplifying yields (4) and (5) in the text.