PHILOSOPHICAL
TRANSACTIONS
OF
THE ROYAL
SOCIETY B

Phil. Trans. R. Soc. B (2008) **363**, 3503–3514
doi:10.1098/rstb.2008.0146
Published online 19 September 2008

Review

Theoretical and empirical evidence for the impact of inductive biases on cultural evolution

Thomas L. Griffiths[1,*], **Michael L. Kalish**[2] **and Stephan Lewandowsky**[3]

[1]*Department of Psychology, University of California, 3210 Tolman Hall No. 1650,
Berkeley, CA 94720-1650, USA*
[2]*Institute of Cognitive Science, University of Louisiana at Lafayette, Lafayette, LA 70501, USA*
[3]*Department of Psychology, University of Western Australia, Perth, WA 6009, Australia*

The question of how much the outcomes of cultural evolution are shaped by the cognitive capacities of human learners has been explored in several disciplines, including psychology, anthropology and linguistics. We address this question through a detailed investigation of transmission chains, in which each person passes information to another along a chain. We review mathematical and empirical evidence that shows that under general conditions, and across experimental paradigms, the information passed along transmission chains will be affected by the inductive biases of the people involved—the constraints on learning and memory, which influence conclusions from limited data. The mathematical analysis considers the case where each person is a rational Bayesian agent. The empirical work consists of behavioural experiments in which human participants are shown to operate in the manner predicted by the Bayesian framework. Specifically, in situations in which each person's response is used to determine the data seen by the next person, people converge on concepts consistent with their inductive biases irrespective of the information seen by the first member of the chain. We then relate the Bayesian analysis of transmission chains to models of biological evolution, clarifying how chains of individuals correspond to population-level models and how selective forces can be incorporated into our models. Taken together, these results indicate how laboratory studies of transmission chains can provide information about the dynamics of cultural evolution and illustrate that inductive biases can have a significant impact on these dynamics.

Keywords: cultural evolution; Bayesian models; learning

1. INTRODUCTION

Much of human knowledge is acquired not by interacting directly with the physical world, but by interacting with other people. The concepts we use, the social conventions we obey and the languages we speak are often learned by observing examples, behaviour or speech produced by other people. These processes of knowledge transmission constitute a basic element of cultural evolution and have been the object of extensive research in psychology (e.g. Bartlett 1932; Mesoudi 2007), anthropology (e.g. Cavalli-Sforza & Feldman 1981; Boyd & Richerson 1985; Sperber 1996) and linguistics (e.g. Kirby 2001; Briscoe 2002; Nowak *et al.* 2002). A key question in all cases is how the minds of human learners shape the outcomes of cultural evolution: how inductive biases—the constraints on learning and memory, which influence our conclusions from limited data—relate to the concepts, conventions and languages which appear in human societies.[1]

In this paper, we explore one part of this question by analysing the effects of inductive biases on one simple form of knowledge transmission: the case where information is passed from one person to another (figure 1). In this case, each person observes data generated by the previous person, forms a hypothesis about the process that produced those data and then uses that hypothesis to generate data for the next person. For example, a language learner might infer the grammar of a language by hearing the utterances of another person, and then use that grammar to generate utterances that are heard by someone else. The languages spoken by the people in this chain will gradually change over time as a consequence of this process. Transmission chains of this kind represent each generation of learners with just one person, and thus do not allow us to explore the influences of individuals within a generation on one another; nonetheless, they provide a powerful tool for exploring how knowledge changes when transmitted across generations.

Our analysis of transmission chains (also known as diffusion chains) uses a mixture of mathematical modelling and laboratory experiments with human participants. Mathematical models are widely used in the study of cultural evolution, often drawing on the rich body of mathematical models of biological evolution (Cavalli-Sforza & Feldman 1981; Boyd & Richerson 1985; Nowak *et al.* 2002). Laboratory experiments are used more rarely, although there exist both classic and

[*] Author for correspondence (tom_griffiths@berkeley.edu).

One contribution of 11 to a Theme Issue 'Cultural transmission and the evolution of human behaviour'.

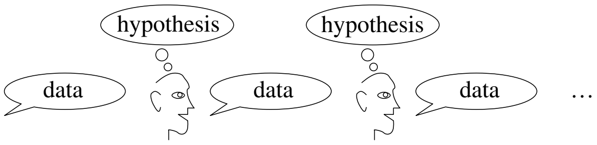

Figure 1. Transmission chains provide a simple setting for studying cultural transmission that has been used in psychology, anthropology and linguistics. In a transmission chain, each agent observes the data generated by the previous agent, forms a hypothesis about the source of these data and then uses that hypothesis to generate data for the next agent.

more recent studies of this kind (see Mesoudi 2007; Caldwell & Millen 2008; Mesoudi & Whiten 2008). Combining mathematical modelling with laboratory experiments gives us the opportunity to test the predictions of our models. Because the mechanisms of cultural evolution are fundamentally psychological, involving processes such as learning, memory and decision-making, using the methods of cognitive psychology allows us to determine whether we have accurately characterized these mechanisms.

We seek to describe how human inductive biases change the information being transmitted. Both learning and remembering involve inductive problems, requiring people to form hypotheses that go beyond the limited data that are available to them (e.g. Anderson 1990). Learning language is a classic example of an inductive problem, with the grammar of the language being underdetermined by the utterances a learner observes. Similar problems arise in other settings, such as determining whether a social convention such as tipping applies based on a few examples or reconstructing a briefly glimpsed experimental stimulus. Inductive biases are the factors that lead a learner to choose one hypothesis over another when both are equally consistent with the observed data. In language learning, such biases might favour languages of certain forms over others, whereas in the case of tipping they might reflect beliefs about social structures. While previous work has explored how relatively simple 'direct biases' that influence whether an agent adopts a hypothesis affect knowledge transmission (Boyd & Richerson 1985), we aim to obtain general results characterizing the consequences of arbitrarily complex inductive biases.

Exploring the effects of inductive biases on knowledge transmission requires having a means of expressing these biases. We do this by analysing transmission chains formed of agents who use Bayesian inference, a mathematical theory that provides a rational solution to inductive problems. Bayesian models make inductive biases explicit and have accounted for human learning (Anderson 1991; Tenenbaum & Griffiths 2001; Griffiths & Tenenbaum 2005) and memory (Anderson & Milson 1989; Shiffrin & Steyvers 1997; Griffiths *et al.* 2007) with considerable success. Examining how knowledge transmission by Bayesian agents is affected by the inductive biases of those agents gives us a very general framework, whose assumptions overlap with accounts of rational behaviour in economics and statistics. This framework makes predictions about the outcomes of cultural evolution, which we can test in the laboratory with human participants.

Our central thesis is that the inductive biases of individuals have a significant effect on the information conveyed along a transmission chain, and that this suggests that inductive biases may play a significant role in cultural evolution more broadly. In support of this thesis, we present a basic mathematical result—that information passed along a transmission chain formed of the Bayesian agents ultimately comes to reflect the inductive biases of those agents (Griffiths & Kalish 2005, 2007; Kirby *et al.* 2007)—and summarize a series of experiments with human participants, which bear out this prediction (Kalish *et al.* 2007; Griffiths *et al.* 2008). We also show that this analysis can be generalized to populations as well as chains of individuals, producing parallels with formal models of biological evolution, and that in such a context the inductive biases of individual learners can have a greater effect on the outcome of cultural evolution than selective forces.

We proceed as follows: §2 reviews the significance of questions about inductive biases and cultural evolution in anthropology, psychology and linguistics; §3 discusses how these different disciplines have converged on the use of transmission chains and summarizes our mathematical analyses; §4 presents empirical results bearing out the predictions of this account; §5 outlines how our approach relates to the models of biological evolution and the relative importance of inductive biases and selective forces in cultural evolution; and §6 presents our conclusions.

2. RELATING INDUCTIVE BIASES AND CULTURAL EVOLUTION

Inductive problems feature prominently in cognition. Questions about how people learn categories, functional relationships or languages ultimately reduce to questions about human inductive biases. Typically, research with adult participants explores the form of these biases, such as what kinds of categories are easy to learn (Shepard *et al.* 1961), whereas researchers in cognitive development seek to understand the origins of those biases (e.g. Spelke *et al.* 1992; Gopnik & Meltzoff 1997). Recently, evolutionary psychologists have suggested that we can obtain answers to these questions by looking at 'human universals' (Brown 1991)—the beliefs and practices which seem to be common to all human societies (e.g. Pinker 2002).

Anthropologists have explicitly explored the relationship between inductive biases and cultural evolution. Sperber (1985, 1996), Boyer (1994, 1998) and Atran (2001, 2002) have argued that processes of cultural transmission provide the opportunity for inductive biases, such as ontological commitments about the kinds of entities that exist, to manifest themselves in culture. This argument is based on the significant role that learning and memory play in cultural transmission. Sperber (1996, p. 84) states that 'the ease with which a particular representation can be memorized' will affect its transmission, and Boyer (1994, 1998) and Atran (2001) emphasize the effects of inductive biases on memory. This idea has some empirical support. For example, Nichols (2004) showed that social conventions based on disgust were more likely to survive several decades of cultural transmission than those without this emotional component. This advantage is consonant with the large body of research showing that emotional events are often remembered better than comparable events that are lacking an emotional component (for a review, see Buchanan 2007).

The role of memory and learning in cultural transmission has also led to arguments against applying mathematical models of biological evolution to cultural evolution (e.g. Cavalli-Sforza & Feldman 1981; Boyd & Richerson 1985), on the grounds that imperfect inferential transmission is very different from the more reliable copying of genes, which underlies biological evolution (Boyer 1998; Atran 2001; Sperber & Claidiére 2006). In particular, cognitive factors that transform knowledge in a way that is analogous to the mutation of genes may play a more significant role in cultural evolution than external selective forces that favour one piece of knowledge over another. Henrich & Boyd (2002) presented several simple models intended to defuse these arguments. For example, one model showed that in the presence of strong 'cognitive attractors' that make agents more likely to adopt particular pieces of knowledge, weak selective forces that increased the value of different knowledge were sufficient to favour one attractor over another as the outcome of cultural evolution. We return to the question of how inductive biases and selection interact in §5.

Research on language evolution also explores the relationship between inductive biases and cultural transmission, examining how constraints on language learning influence the languages that a population of learners comes to speak. Human languages form a subset of all logically possible communication schemes, with some properties being shared by all languages (Greenberg 1963; Comrie 1981; Hawkins 1988). Traditionally, these 'linguistic universals' are explained by appealing to the constraints of an innate system specific to the acquisition of language (e.g. Chomsky 1965). A popular alternative explanation is that the universal properties of human languages have arisen as a consequence of languages being learned anew by each generation, with each learner having only weak, domain-general inductive biases (e.g. Kirby 2001). This alternative explanation relies upon the possibility that cultural transmission can emphasize the inductive biases of language learners, allowing such weak biases to be translated into strong and systematic universals of the kind seen in human languages.

The effects of cultural transmission on languages have also been the subject of extensive observational and experimental analysis. Creolization, the formation of a more regular system of communication from a piecemeal pidgin, has traditionally been one of the strongest arguments for constraints on language acquisition influencing the structure of languages (Bickerton 1981), and typically occurs when a language is passed from one generation to the next. Experiments investigating how adults and children learn artificial but realistic languages have provided support for the idea that language learning by children plays an important role in this process, showing that children tend to regularize probabilistic elements of languages (making them more deterministic) to a greater extent than adults (Hudson-Kam & Newport 2005). Recent work has also explored how languages are formed and change across generations through the observation of the development and transmission of sign languages (Senghas et al. 2004), complementing an extensive theoretical and empirical literature on language creation and change (DeGraff 1999).

The preceding examples illustrate that all the three disciplines discussed—psychology, anthropology and linguistics—could be informed by a deeper understanding of how inductive biases affect knowledge transmission.

3. USING TRANSMISSION CHAINS TO MODEL CULTURAL EVOLUTION

In addition to sharing common questions about the influence of inductive biases on cultural transmission, psychologists, anthropologists and linguists have all used a common paradigm to explore these questions, examining what happens when information is transmitted along a single chain of individuals, as illustrated in figure 1. Such transmission chains provide a way to study one of the basic elements of cultural evolution—how information changes when passed from one person to another—in isolation, making it possible to study it in detail. While this analysis ignores many of the other factors that are important to the creation and change of concepts and languages, such as interactions between individuals within a generation (Steels 2003; Galantucci 2005; Garrod et al. 2007), understanding how each of these factors operates in isolation will ultimately help understand their combination.

The use of transmission chains in psychology was pioneered by Bartlett's (1932) 'serial reproduction' experiments, in which participants were shown a stimulus and then asked to reproduce it from memory, with their recalled version being presented to the next participant and so on. Bartlett argued that reproductions seem to become more consistent with the cultural biases of the participants as the number of successive reproductions increases. However, these arguments were largely anecdotal and lacked quantitative rigor. Nonetheless, serial reproduction has become one of the primary methods that psychologists have used to explore the effects of cultural transmission, and similar experiments are used by anthropologists and biologists to examine what kinds of cultural concepts persist over time and whether non-human animals can transmit information across generations (for a review, see Mesoudi 2007; Mesoudi & Whiten 2008; Whiten & Mesoudi 2008).

In linguistics, the study of transmission chains has largely been restricted to *simulations* of the process of language change. In these 'iterated learning' simulations, a sequence of agents each learns a language by observing the utterances of the previous agent, and then in turn produces utterances that are observed by the next agent (Kirby 2001; see Smith & Kirby 2008). Simulations have shown that languages with interesting structure emerge from iterated learning with a variety of learning algorithms (Kirby 2001; Brighton 2002; Smith et al. 2003). In particular, basic properties of human languages such as compositionality—the use of different parts of an utterance to describe different aspects of an event—can be produced by very simple learning algorithms, without requiring innate language-specific constraints on learning (e.g. Smith *et al.* 2003).

The prevalence of transmission chains in research on cultural evolution is due in part to their simplicity as a model of knowledge transmission. This simplicity also makes transmission chains amenable to mathematical analysis. In the remainder of this section, we summarize the behaviour of transmission chains consisting of a sequence of the Bayesian agents (Griffiths & Kalish 2005, 2007; Kirby *et al.* 2007).

(a) *Chains of Bayesian agents*

Following the schema shown in figure 1, we have a sequence of agents, each of whom observes data d and forms a hypothesis h about the knowledge of the previous agent responsible for generating those data. What form the data and hypotheses take will depend on the kind of knowledge being transmitted: for concepts, data could be instances of that concept and hypotheses rules that characterize it; for social conventions, data could be observations of the behaviour of others and hypotheses the circumstances under which a convention applies; and for languages, data could be a set of utterances and hypotheses grammars. We assume that each learner selects a hypothesis by sampling from a distribution $P_{\mathrm{LA}}(h|d)$, where LA refers to some learning algorithm, and generates data by sampling from a distribution $P_{\mathrm{PA}}(h|d)$, where PA refers to some production algorithm. Using h_n and d_n to represent the hypothesis formed and the data generated by the nth learner, respectively, this defines a stochastic process on (h_n, d_n) pairs.

A first observation is that this process is a Markov chain: a sequence of random variables in which each variable depends only on that which precedes it. In our case, the hypothesis–data pair (h_n, d_n) is independent of all preceding pairs given (h_{n-1}, d_{n-1}). Marginalizing out (i.e. summing over) either hypotheses or data makes it possible to define Markov chains on just d_n or h_n, respectively. It is often particularly convenient to study the Markov chain on hypotheses. If the number of hypotheses is finite, the probability of the nth learner adopting hypothesis i given that the $n-1$th learner held hypothesis j is given by the transition matrix $\boldsymbol{Q}$, with entries

$$\begin{aligned} q_{ij} &= P(h_n = i | h_{n-1} = j) \\ &= \sum_d P_{\mathrm{LA}}(h_n = i|d) P_{\mathrm{PA}}(d|h_{n-1} = j), \end{aligned} \tag{3.1}$$

which will depend on the learning and production algorithms adopted by the learners.

Reducing the process of cultural transmission to a Markov chain makes it easy to ask questions about the outcome of such a process. Provided the Markov chain satisfies a set of easily checked conditions, it will converge asymptotically to a *stationary distribution* (Norris 1997). In the case of the Markov chain on hypotheses identified above, this means that the probability that the nth learner entertains a particular hypothesis will converge to a fixed value as n becomes large, regardless of the hypothesis entertained by the first learner. Determining the consequences of using a particular learning algorithm is thus a matter of determining how that learning algorithm influences the stationary distribution. This distribution can be found numerically by computing the first eigenvector of the transition matrix (such as the matrix $\boldsymbol{Q}$ defined in equation (3.1), but in some cases it is also possible to give an analytic characterization.

Transmission chains formed of the Bayesian agents provide one case in which an analytic stationary distribution can be obtained. If we use a probability distribution over hypotheses $P(h)$ to encode an agent's degrees of belief in each hypothesis before seeing the data (known as the *prior* distribution), the corresponding distribution $P(h|d)$ after seeing the data d (known as the *posterior* distribution) is obtained by applying Bayes' rule

$$P(h|d) = \frac{P(d|h)P(h)}{\sum_{h' \in \mathcal{H}} P(d|h')P(h')}, \tag{3.2}$$

where $P(d|h)$ (known as the likelihood) is the probability of seeing the particular data d if the particular hypothesis h is true, and the sum in the denominator ranges over the set of all possible hypotheses, $\mathcal{H}$. The Bayesian inference provides a useful framework for exploring questions about inductive biases, since the prior $P(h)$ effectively encodes the inductive biases of the agent, being a source of additional information or constraints that discriminate between hypotheses with equal likelihoods. Thus, hypotheses with lower prior probability are harder to learn or remember, requiring stronger evidence to achieve high posterior probability.

The assumption that the agents use Bayesian inference reduces the psychological complexities of learning to a single equation. At first glance, this might appear to ignore a long tradition of work on understanding human learning by cognitive psychologists; however, rather than ignoring that knowledge, our approach merely characterizes human learning at a higher level of abstraction, often referred to as the 'computational level' (Marr 1982). That is, we are exclusively concerned with the *outcome* of learning but have no commitment to a specific *process* by which it occurs. Many available models of learning and skill acquisition may provide helpful process instantiations of the Bayesian agents in our computational level of description, and formal equivalences exist between some of these process models and Bayesian inference (e.g. Ashby & Alfonso-Reese 1995).

The learning algorithms we will consider are based on the posterior distribution produced by applying Bayes' rule. 'Learning' in the present context refers to the choice of a hypothesis about the data, so perhaps the simplest algorithm is to sample a hypothesis from the posterior. Using this algorithm, the distribution $P_{\mathrm{LA}}(h|d)$ becomes

$$P_{\mathrm{samp}}(h|d) = \frac{P_{\mathrm{PA}}(d|h)P(h)}{\sum_{h' \in \mathcal{H}} P_{\mathrm{PA}}(d|h')P(h')}, \tag{3.3}$$

where we place no constraints on the production algorithm PA, but assume that the learning algorithm employed by the agents draws on accurate knowledge of this distribution.[2]

With these specific assumptions about the form of the learning algorithm in hand, we are able to analyse the stationary distribution of the resulting Markov chain. Griffiths & Kalish (2005) showed that the stationary distribution of the Markov chain on hypotheses is the prior distribution, $P(h)$. A more extensive analysis performed by Griffiths & Kalish (2007) also provided stationary distributions for Markov chains on data and hypothesis–data pairs, and pointed out a correspondence between the latter and a Markov chain Monte Carlo algorithm called Gibbs sampling (Geman & Geman 1984), commonly used in Bayesian statistics. In a nutshell, these mathematical results imply that irrespective of the stimuli presented at the outset, the final result of iterated learning across generations is the expression of the learners' inductive biases.

Convergence to the prior provides a simple answer to the question of how the inductive biases of individuals affect the outcome of cultural evolution. It indicates that the probability that a particular hypothesis—a language, religious concept or social norm—will emerge as the result of being transmitted from one person to another is simply the prior probability of that hypothesis. This means that inductive biases—the constraints on learning that characterize the minds of individuals—will lie in a direct one-to-one correspondence with the outcomes of knowledge transmission. Returning to the various claims about cultural evolution made above, this analysis is consistent with Bartlett's conclusions about serial reproduction revealing cultural biases, with the arguments of Boyer (1994, 1998), Sperber (1996) and Atran (2001) concerning the role of human cognition in shaping the information being transmitted, and with the analysis of linguistic universals as the direct outcome of constraints on language acquisition.[3]

Making what might seem like a small change to the assumptions about the learning algorithm used by our Bayesian agents has significant consequences. An alternative to sampling from the posterior distribution is to choose the hypothesis that has the highest posterior probability (known as *maximum a posteriori* or MAP estimation). In this case, the probability of selecting a particular hypothesis becomes

$$P_{\mathrm{MAP}}(h|d) \propto \begin{cases} 1, & h \text{ maximizes } P(h|d), \\ 0, & \text{otherwise}, \end{cases} \tag{3.4}$$

where $P(h|d)$ is computed as in equation (3.3), and the constant of proportionality is determined by the number of maxima of $P(h|d)$. Griffiths & Kalish (2007) showed that in this case a small difference in the prior $P(h)$ can result in a big difference in the stationary probability of a hypothesis. Kirby *et al.* (2007) showed that moving from sampling to MAP estimation increases the magnitude of the effect of the prior on the outcome of knowledge transmission, with hypotheses that are slightly favoured by the prior being over-represented in the stationary distribution. These results paint a slightly different picture of the relationship between inductive biases and cultural universals, showing that weak inductive biases can be magnified by the process of cultural transmission to produce strong universals. This is still consistent with the claims of psychologists and anthropologists about the importance of cognitive factors in cultural evolution. However, it undermines the inference from cultural universals to equivalently strong constraints on learning, which is part of the traditional interpretation of linguistic universals: if weak biases can be magnified by cultural evolution, then we no longer need to postulate strong constraints to account for the consistency observed in human languages.

(b) *A simple example: two hypotheses*

We illustrate the dynamics of the Bayesian transmission chains with a simple example. In this example, we assume that agents choose between two hypotheses by sampling from their posterior distributions. A similar example covering both sampling and MAP estimation is analysed in detail by Griffiths & Kalish (2007).

The case of two hypotheses naturally maps onto a variety of simple pieces of knowledge that might be transmitted across generations, such as whether the verb in a sentence precedes the object, a certain class of foods is considered sacred or to tip taxi drivers. Inductive biases from a variety of sources, from innate constraints on language learning to the social perception of tipping, could influence the transmission of this knowledge. Using numbers to denote hypotheses, we can summarize the prior distribution over these hypotheses by using π to designate $P(h=1)$. Each agent in a chain has the opportunity to observe a piece of data generated by the previous agent, such as a set of utterances, a labelling of sacred objects or some tipping behaviour. To simplify, we will assume that this piece of data can also take on two values and that these values are indicative of the hypothesis entertained by the agent. This can be done by taking $P(d=k|h=k)=1-\epsilon$ for $k\in\{1,2\}$, where ϵ is a parameter indicating the amount of noise in transmission.

These assumptions provide us with all the information we need to compute the transition matrix of the Markov chain on hypotheses. The prior and likelihood specified by π and ϵ can be substituted into equation (3.2) to give the posterior distributions,

$$P(h=1|d=1) = \frac{(1-\epsilon)\pi}{(1-\epsilon)\pi + \epsilon(1-\pi)},$$

$$P(h=1|d=2) = \frac{\epsilon\pi}{\epsilon\pi + (1-\epsilon)(1-\pi)},$$

where the probabilities for $h=2$ are obtained from the fact that the posterior sums to 1. Substitution into equation (3.1) can be used to compute the transition matrix, summing over the values $d\in\{1,2\}$. Since probabilities sum to 1, we need to specify only two of the entries of $\boldsymbol{Q}$, such as q_{12} and q_{21}, to give the full transition matrix. An elementary calculation yields

$$q_{12} = c\pi \qquad q_{21} = c(1-\pi), \tag{3.5}$$

where $c=\epsilon(1-\epsilon)(1/(1-\epsilon-\pi+2\epsilon\pi)+1/(\epsilon+\pi-2\epsilon\pi))$. This indicates that the probability of moving from hypothesis 2 to 1 is proportional to the prior probability π, but the constant of proportionality is strongly influenced by the noise rate ϵ, increasing as ϵ increases.

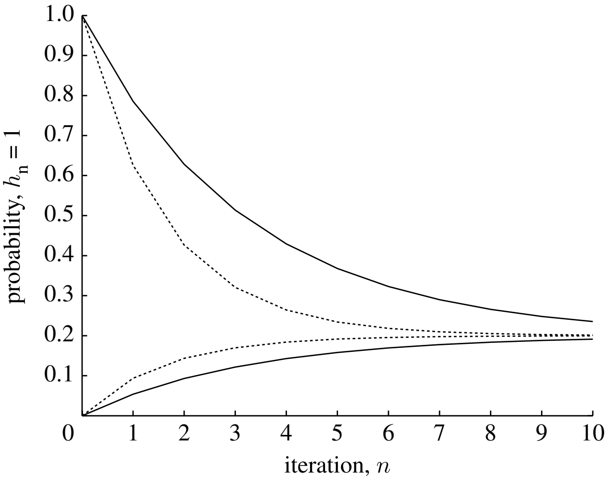

Figure 2. Dynamics of the probability of an agent adopting hypothesis 1 as a function of the number of generations of transmission. As the number of generations increases, the probability of choosing h_1 converges to the prior probability, $\pi = 0.2$. The noise parameter ϵ determines the rate of convergence, with $\epsilon = 0.01$ (solid lines) converging more slowly than $\epsilon = 0.05$ (dotted lines).

The transition matrix can be used to characterize the dynamics and asymptotic consequences of cultural transmission. The probability that an agent chooses a particular hypothesis after n iterations is given by $\boldsymbol{Q}^n\boldsymbol{p}$, where $\boldsymbol{p}$ is a vector specifying the distribution over hypotheses used to generate the first piece of data. Figure 2 shows how this quantity evolves over time for $\pi = 0.2$ and $\epsilon \in \{0.01, 0.05\}$. Regardless of whether the first piece of data was generated from hypothesis 1 or 2, just 10 iterations are sufficient to bring the probability that an agent selects a hypothesis close to the prior probability π. Increasing the value of ϵ (and hence the noise in the transmission) increases the rate of convergence, making it easier for an agent to entertain a hypothesis different from that of the previous agent.

The first eigenvector of $\boldsymbol{Q}$ is a vector $\boldsymbol{\theta}$ such that $\boldsymbol{Q\theta} = \boldsymbol{\theta}$. It makes intuitive sense that this should be the stationary distribution of the Markov chain, since this defines a distribution that does not change through further application of the stochastic process defined by $\boldsymbol{Q}$ (i.e. by definition of eigenvectors, $\boldsymbol{Q}^n\boldsymbol{\theta} = \boldsymbol{\theta}$ for all n). Since $\theta_2 = 1 - \theta_1$, we can reduce this definition to an equation in a single variable,

$$(1 - q_{21})\theta_1 + q_{12}(1 - \theta_1) = \theta_1, \tag{3.6}$$

which has the solution $\theta_1 = q_{12}/(q_{12} + q_{21})$. Substituting the values for q_{12} and q_{21} from equation (3.5) into this solution, we obtain $\theta_1 = \pi$. This indicates that the stationary probability of hypothesis 1 is π, being equal to its prior probability and consistent with the convergence shown in figure 2.

(c) *Summary*

Transmission chains provide a simple way to study one of the basic forces in cultural evolution—the way that knowledge changes when transmitted from person to person. This simplicity is paralleled in the mathematical analysis of such systems that reduce to Markov chains. When the chain is composed of Bayesian agents, we can make precise predictions about the effects of inductive biases (expressed in the priors of those agents) on knowledge transmission: the probability that an agent considers a hypothesis will converge to the prior probability of that hypothesis. We next examine whether these predictions are borne out in the laboratory.

4. SIMULATING CULTURAL EVOLUTION IN THE LABORATORY

Empirical tests of the idea that transmission chains converge to the agents' prior distributions face two obstacles. First, we must know what the priors are, so that we can recognize how closely they are approximated by the stationary distribution. Second, we must be able to determine when (and if) a chain has converged. The first constraint led us to consider two simple tasks for which previous research provided strong evidence as to the general structure of the prior. The second constraint led us to a design that employed multiple chains starting from different initial states. Convergence has occurred when all chains produce similar results despite their diverse initial conditions.

(a) *Learning categories*

The simplest example of this method, and perhaps the best instance of a known prior in an appropriate domain, is a study in which people learned to extend a partially specified category to a set of novel items (Griffiths *et al.* 2008, Experiment 1B). The items all varied on three binary dimensions and the categories divided the eight items into two classes of four. If we do not distinguish structures that differ only in the assignment of physical features to the binary dimensions, there are only six types of such categories (figure 3*a*). To illustrate, if the three binary dimensions defined geometric objects by shape (e.g. circle or square), size (e.g. small or large) and colour (e.g. black or white), then a type I category might differentiate all squares (regardless of size or colour) from the circles, whereas a type II category might pick out white squares and differentiate them from black circles (regardless of size).

Psychological research has told us a good deal about how people learn these categories. In particular, Shepard *et al.* (1961) showed that the six types of categories have a canonical ordering of difficulty, with types I and II being significantly easier to learn than the others. The robustness of this finding (e.g. Nosofsky *et al.* 1994) suggests that it is an effective index of the prior over the six category types: the more difficult a category is, the more data it requires to learn and hence the lower its prior probability.

We used these category types to explore whether people's inductive biases—reflected in the difficulty-of-learning results—would influence the outcome of cultural transmission. Our stimuli were 'amoebae' whose nuclei varied along the three dimensions of shape, size and colour mentioned above (after Feldman 2000). People were asked to make inferences about 'species' of amoebae based on examples. On each trial of the experiment, a participant was shown three amoebae that were stated to belong to a species, and asked to identify the fourth amoeba belonging to that species. To do so, all possible four-item categories that

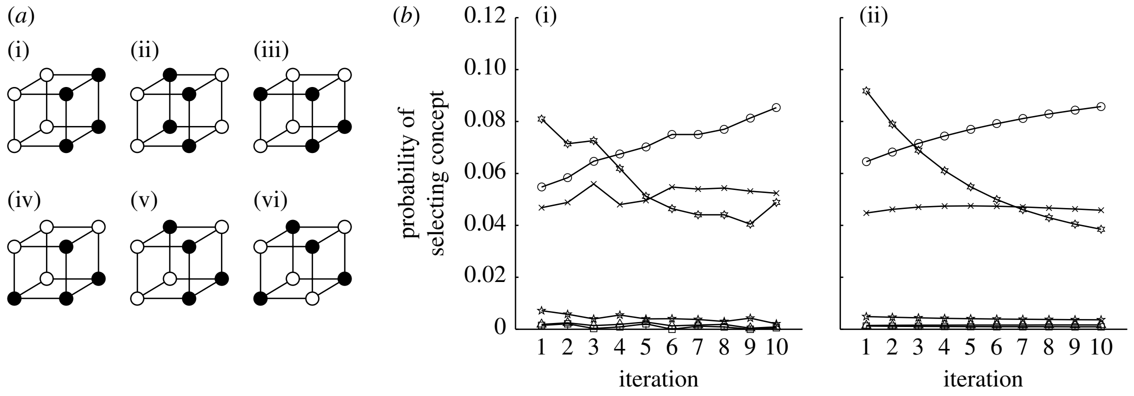

Figure 3. Transmission chains for categories. (*a*) If we consider categories that are sets of four objects defined on three binary dimensions and ignore the assignment of the dimensions to the physical properties of those objects, there are just six possible category types (i–vi), types I–VI (Shepard *et al.* 1961). Each of the six types is illustrated on a cube, where each dimension of the cube corresponds to one of the binary dimensions and the vertices are the eight objects. Filled circles represent members of an example category of that type. Type I categories are defined on one dimension; type II uses two dimensions; types III, IV and V are one dimension plus an exception and type VI uses all three dimensions. (*b*) Transmission chains were constructed by showing people three objects drawn from a category and asking them to indicate, from a set of possible alternatives, which object completed the set. The objects seen by the next person were selected at random from the set selected by the previous person. The probability with which people selected categories of the six types changes as a function of the number of generations of a transmission chain, as predicted by a Bayesian model using a prior estimated from human learning data. In particular, the probabilities of types I and VI increase and decrease, respectively. (i) Human participants and (ii) Bayesian model (circles, type I; crosses, type II; triangles, type III; squares, type IV; five-point stars, type V; six-point stars, type VI). Further details are provided in Griffiths *et al.* (2008).

contained the three original amoebae and one other amoeba were presented to the participant who selected the category deemed most likely. Formally, the three original amoebae are the data d and the response alternatives are the hypotheses h. Participants were implicitly being asked to compute $p(h|d)$ and use it to select one of the alternatives.

Each of the participants in the experiment completed a series of trials, of which a subset were linked to the responses of other people via transmission chains. Specifically, the participants were randomly grouped into seven ‘families’ of 10 generations each, with responses transmitted between members of each family. For the first participant in each family, the amoebae seen on each trial were sampled uniformly at random from the set of four matching a category structure of one of the six types, with the six types appearing with equal probability. The amoebae seen by the next participant were then sampled from the set of four selected by the first participant and so forth.

Under the mathematical analysis presented above, the frequency of each category type in each generation should come to approximate the prior as the number of generations increases. This is precisely what was observed empirically: the frequency of type I concepts increased and type VI decreased over the course of the experiment, and types I and II dominated responses by the end of the experiment (figure 3*b*). The use of a finite hypothesis space made it possible to compute a full transition matrix for this Markov chain, and the numerical predictions of the resulting Bayesian model were strongly consistent with the observed data (figure 3*b*).

(b) *Learning functions*

In contrast to the limited set of hypotheses available to learners with the concepts described above, most inductive problems allow for a vast number of hypotheses. One such task is function learning, where a metric stimulus value (such as the dosage of a drug or driving speed) is related to a metric criterion (such as the response to the drug or stopping distance). Such relationships can have arbitrary complexity, but people nonetheless appear to have strong priors over the space of possible relationships. Kalish *et al.* (2004), in reviewing the literature on function learning, observed that people generally assume (and are the quickest to learn) increasing linear functions where the criterion increases in direct proportion to the stimulus. This is consistent with an inductive bias that favours such functions.

Exploiting knowledge about human inductive biases for this task, Kalish *et al.* (2007) conducted an experiment in which people formed a transmission chain for function concepts. In this experiment, each generation of participants received 50 trials of training on a single function. On each trial, the value of the stimulus was presented as a visual magnitude, being the width of a horizontal bar on a computer screen. Participants responded by adjusting the height of a vertical bar and then received corrective feedback (by displaying the correct magnitude next to the response bar). After training, participants responded to 100 stimuli that covered the entire possible range of magnitudes without receiving feedback.

As in the experiment described above, the data seen by the participants were influenced by the responses of other participants. Participants were arranged into eight families of nine generations, for each of four conditions. The conditions differed with respect to the function used to generate the training data seen by the first generation of participants: those initial values were drawn either from a positive linear, negative linear or quadratic function, or entirely at random. For example, a participant trained on the negative linear function would see a series of training pairings where large stimulus values (i.e. long bars) were paired with small

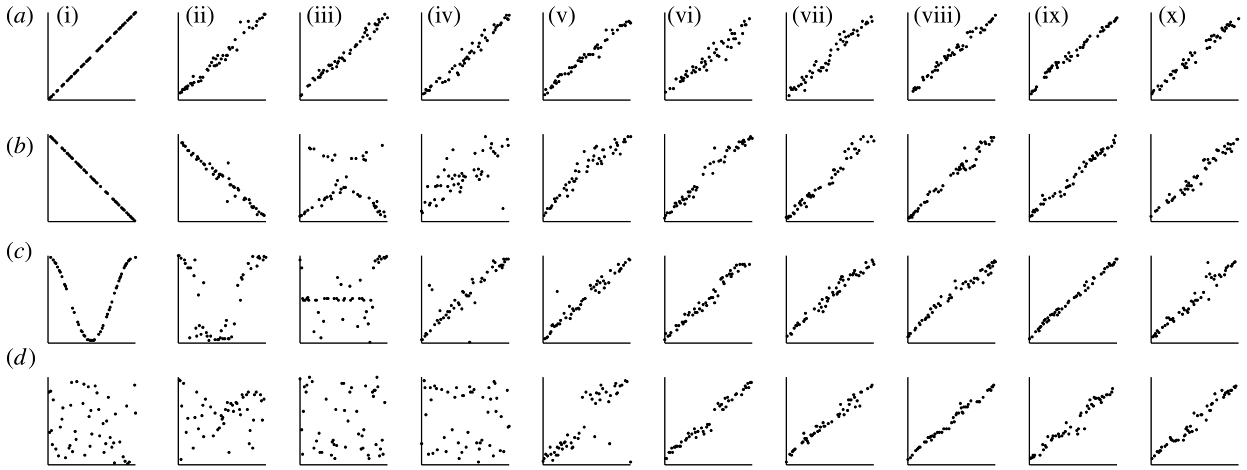

Figure 4. Representative results for transmission chains with human participants in which people learn functions. (*a*–*d*) Each row shows a single chain. (i) The (x, y) pairs were presented to the first participant in the chain, being represented as the width and height of horizontal and vertical rectangles, respectively. Participants then made predictions of the value of y for new x values ((ii) $n=1$, (iii) $n=2$, (iv) $n=3$, (v) $n=4$, (vi) $n=5$, (vii) $n=6$, (viii) $n=7$, (ix) $n=8$, (x) $n=9$). These predictions formed the (x, y) pairs given to the next person in the chain, whose data appear in (ii)–(x) and so forth. Consistent with the previous research exploring human inductive biases for function learning, chains produced linear functions with mostly positive slopes, regardless of whether they were initialized with (*a*) a positive linear function, (*b*) a negative linear function, (*c*) a nonlinear function or (*d*) a random collection of points.

criterion values (i.e. short bars) and vice versa. The responses of each participant on 50 of the test trials were taken as the data used to train the participant in the next generation of that family.

Representative families from the four conditions are shown in figure 4. Two features of the data from these chains are immediately apparent. First, striking changes in the stimulus–criterion functions across generations were observed, but only sporadically. This indicates that people's acquired functions were generally very easy for the next generation to learn. Second, notwithstanding the dramatic differences between functions at the outset, across generations all of the initial functions gradually disappeared and transited into only one of two stable functions: positive linear (28 out of 32 families) and negative linear (4 out of 32), both with approximately unit slope. These results are consistent with the previous work suggesting that people's priors are centred on positive and negative linear functions and they support the predictions of our formal analysis.

(c) *Summary*

Laboratory experiments involving transmission chains for concepts that have been extensively studied by psychologists provide a direct test of the predictions of our formal framework. By using categories and functions—concepts for which human inductive biases are well understood—we were able to investigate whether these biases influence the outcome of knowledge transmission. The results support the conclusion that knowledge transmission converges to an equilibrium determined by the inductive biases of learners, with categories and functions that people find easier to learn becoming more prevalent across generations. Flynn (2008) reports an analogous result with small children, who very quickly discard irrelevant information when transmitting a sequence of problem-solving moves to an observer in the next generation.

Our laboratory results have implication for views of human cultural evolution. In particular, the data are consonant with the view that cultural representations tend to be 'recurrent'—that is, many aspects of culture transcend beyond isolated times and places (e.g. Boyer 1998). Our repeated demonstrations that inductive biases determine the final outcome of knowledge transmission provide an empirical foundation for claims by anthropologists and psychologists that human cognitive capacities will influence the ideas that appear in human societies, such as Boyer's (1998) claim that religious concepts are influenced by people's 'intuitive ontologies'—i.e. the distinctions they draw between classes of objects from a very early age.

5. RELATING CULTURAL AND BIOLOGICAL EVOLUTION

We next consider some connections between the theoretical and empirical analyses presented thus far and mathematical models of biological evolution. These connections generalize our results beyond the simple case of transmission chains. Mathematical models of biological evolution are often applied to cultural evolution (Cavalli-Sforza & Feldman 1981; Boyd & Richerson 1985), and it is common to see both informal (Deacon 1997; Kirby 1999) and formal (Nowak *et al.* 2002) analogies between languages and genotypes as objects of evolution. We first discuss how our results relate to standard analyses of evolutionary dynamics, by showing that the evolution of population proportions in the absence of selection is intimately related to the behaviour of transmission chains. We then discuss what this connection tells us about the role of selection in cultural evolution.

(a) *Transmission chains and the replicator dynamics*

The basic model of deterministic evolution is based on the replicator dynamics (e.g. Hofbauer & Sigmund

1998). Let x_i denote the proportion of a population of agents entertaining hypothesis i at a given moment t, and q_{ij} denote the probability that a learner chooses hypothesis i after seeing the data generated from hypothesis j, as defined in equation (3.1). If we assume that each learner learns from a random member of the population, then the population proportions evolve as

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = \sum_j f_j q_{ij} x_j - \phi x_i, \tag{5.1}$$

where f_j is the *fitness* of people who subscribe to hypothesis j; $\phi = \sum_k f_k x_k$ is the mean fitness; and the second term on the right-hand side ensures that $\sum_i x_i = 1$. In biological evolution, fitness reflects the number of offspring produced by an individual of a particular type. In cultural evolution, it is more natural to interpret fitness as influencing the probability with which an individual chooses an agent from the previous generation as a source of data. If agents are selected with probability determined by f_j, the same dynamics hold.[4]

Equation (5.1) has been extensively applied to cultural evolution for the case of languages, in the form of the 'language dynamical equation' explored by Nowak *et al.* (2001, 2002). In this work, fitness is typically assumed to be a function of how well speakers of a particular language can communicate with the population at large, implementing a selection pressure for communication. If we instead assume that all speakers have equal fitness, $f_j = 1$, equation (5.1) simplifies to

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = \sum_j q_{ij} x_j - x_i, \tag{5.2}$$

which is a linear dynamical system. This is a 'neutral' model, in which there are no selective forces favouring one language or hypothesis over another. A special case of this model was analysed by Komarova & Nowak (2003).

The neutral model characterizes the evolution of a population in the absence of selection, and thus provides a valuable null hypothesis against which to evaluate claims about selective forces, as well as a way to study the effects of mutation. It also gives us a way to connect the replicator dynamics to transmission chains. The asymptotic behaviour of this linear dynamical system is straightforward to analyse: it converges towards an equilibrium at the first eigenvector of the transition matrix $\boldsymbol{Q}$ (for details, see Griffiths & Kalish 2007). This means that the neutral form of the replicator dynamics displays asymptotic behaviour that is very similar to that of transmission chains involving discrete generations of single learners. The key difference is in the nature of the quantities that converge: with discrete generations of single learners, it is the probability with which a particular learner entertains hypothesis i that converges to the stationary probability; under the replicator dynamics, it is the *proportion* of the population that entertains hypothesis i that converges to this probability.

The results from the previous sections characterize the consequences of cultural evolution not only for individuals but also for populations. This provides an additional justification for the use of transmission chains in studying cultural evolution: the parallel between the stationary distributions of such chains and the equilibria of the replicator dynamics in populations provides a way to gather clues about the behaviour of populations using a paradigm that is easily simulated in the laboratory.

(b) *Inductive biases can overwhelm selective pressures*

In addition to indicating how transmission chains can inform the study of cultural evolution more broadly, this connection provides us with a way to generalize our mathematical results to cases where selective forces also influence the adoption of hypotheses. This can allow us to evaluate whether inductive biases can play a more significant role in cultural evolution than selection, as suggested by Sperber (1996), Boyer (1998) and Atran (2001), or whether selection is the more powerful force, as argued by Henrich & Boyd (2002). While obtaining general analytical results is difficult, we can at least gain an idea of how these forces interact by returning to our example with just two hypotheses.

With the two hypotheses, the fact that $x_1 + x_2 = 1$ means that we can work with just one variable. We will use x_1, the proportion of agents choosing hypothesis 1, and denote this x for simplicity. In §3*b*, we defined the matrix $\boldsymbol{Q}$ as a function of the prior probability of hypothesis 1, π, and the noise rate, ϵ. In the neutral model from §5*a*, where the fitness of both hypotheses is equal (i.e. each generation chooses an agent to learn from at random from the previous generation with uniform probabilities), the equilibrium of the system is given by finding a value of x such that equation (5.2) is equal to zero. It is straightforward to show that this is equivalent to solving equation (3.6), and thus the equilibrium is given by $x = \pi$. The critical question is how this equilibrium is affected by selection, as represented by unequal fitness for the two hypotheses.

We will assume that the fitness of hypothesis 1 is $f_1 = s$ and hypothesis 2 has constant fitness $f_2 = 1$. We are interested in the case where $s > 1$. This higher fitness might reflect higher social status accorded to those who adopt the hypothesis, greater success in solving problems posed by the environment as a consequence of having this belief or some other indicator of success that might make others more likely to try to learn from these 'fit' individuals. The equilibrium of the resulting system is given by finding x such that equation (5.1) is equal to zero. Simplification for the case of the two hypotheses reduces this to the quadratic equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = (1-s)x^2 + ((1-q_{21})s - q_{12} - 1)x + q_{12}, \tag{5.3}$$

which can be solved by standard methods to find an equilibrium for a particular choice of s, q_{12} and q_{21}. Figure 5*a* shows how the equilibrium changes as a function of s for $\pi = 0.2$ and $\epsilon \in 0.01, 0.05$. As might be expected, increasing s increases the representation of hypothesis 1 in the equilibrium solution.

We can use equation (5.3) to explore the relative contributions of the prior probability of a hypothesis π and the strength of selection s on the equilibrium of this

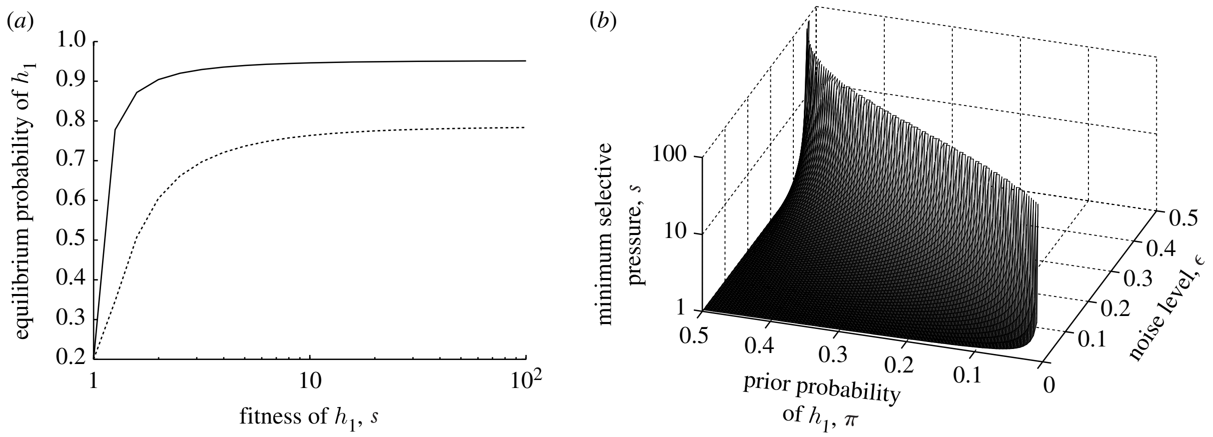

Figure 5. The interaction of selection with inductive biases. (*a*) Increasing the selective pressure in favour of hypothesis 1 increases the representation of that hypothesis in the population. The equilibrium probability of hypothesis 1 for $\pi = 0.2$, $\epsilon \in \{0.01, 0.05\}$ (solid line, dotted line, respectively), and a range of values of the selective pressure s are shown. (*b*) Threshold on s for hypothesis 1 to obtain an equilibrium probability greater than 0.5 as a function of π and ϵ. For values of π and ϵ such that $q_{21} > 0.5$, no value of s produces an equilibrium favouring hypothesis 1.

system. When $s = 1$, we know that the equilibrium value of x will be π. If π is less than 0.5, the equilibrium will be biased against h_1. We might thus ask how large s will have to be in order to overcome this bias, making the equilibrium value of x greater than 0.5. The functions shown in figure 5*a* indicate that this happens relatively quickly for the values of π and ϵ considered above, with the equilibrium passing 0.5 for values of s not much greater than 1. In appendix A, we show that the threshold value of s is

$$s = \frac{1 - 2q_{12}}{1 - 2q_{21}}, \tag{5.4}$$

provided $q_{12} < q_{21} < 0.5$. The first part of this condition follows automatically from the fact that $\pi < (1 - \pi)$, but the second part is more interesting. If $q_{21} > 0.5$, then there is *no* value of s such that the equilibrium favours hypothesis 1. Intuitively, if more than half the agents learning from endorsers of hypothesis 1 adopt hypothesis 2, there is no way that increasing the fitness of hypothesis 1 can push the equilibrium past 0.5.

The requirement that q_{21} be less than 0.5 places strong constraints on the values of π and ϵ, which can support equilibria favouring hypothesis 1. Figure 5*b* shows how the threshold on s behaves as a function of π and ϵ. The threshold rapidly increases as π and ϵ approach values that make q_{21} close to 0.5, and any value of π less than 0.5 has some value of ϵ for which no amount of selection will yield an equilibrium favouring hypothesis 1. For example, $\pi = 0.2$ results in reasonable thresholds on s for small values of ϵ of the kind used in the examples above, but taking $\epsilon = 0.16$ allows the prior to have a sufficiently strong influence on the inferences of the agents that no amount of selection can overcome it. These results thus illustrate how inductive biases can lead a population to an equilibrium that reflects those biases, even if there are other social or environmental factors that strongly favour a different outcome.

6. CONCLUSION

At the start of this paper, we asked a very general question concerning cultural evolution, namely how people's inductive biases (their knowledge and expectations) affect the transmission of languages and concepts. We analysed this general question in the more circumscribed context of transmission chains, in which knowledge is passed from one person to the next. Within this paradigm, the general question about inductive biases becomes the question of how these biases change the information being transmitted. We provided two converging answers: one based on an abstract mathematical analysis and the other based on evidence from behavioural experiments. Both answers suggest that in many circumstances, transmission chains converge to an equilibrium that reflects people's inductive biases.

The mathematical results we summarized apply to learning algorithms based on the Bayesian inference in which observed data are combined with inductive biases expressed as a prior distribution over hypotheses. In this case, the probability with which a person at the end of a transmission chain selects a particular hypothesis converges to a distribution determined by the prior. The data from several experiments were found to be in accord with this prediction: after transmission across a fairly small number of generations, people's responses approximated their known inductive biases in terms of the proportions with which they chose competing hypotheses, for both categorical concepts and continuous functions. In both cases, people's biases were established independently through previous experiments, and, with categorical concepts, direct measurement within the same experiment. The fact that the products of our transmission chains were consistent with these inductive biases suggests that the way people behave in these tasks is sufficiently similar to the Bayesian inference to permit the conclusion that our mathematical results accurately characterize the dynamics of cultural transmission.

These mathematical analyses and experimental results imply two strong statements about cultural evolution in general. First, they indicate that the power of inductive biases can trump the potential stabilization provided by faithful learning. Recall that in the function learning experiment of Kalish *et al.* (2007), the first generation of learners was presented with widely different functions, ranging from positive linear to quadratic and entirely random—nonetheless, after only four or five generations, those different starting points had been absorbed and responses converged to a function that remained stable across further generations. Learning from the data produced by the previous participant was thus insufficient to guarantee faithful cultural transmission, with the influence of inductive biases accumulating with each generation. Second, the analyses reported in §5 suggest that prior biases may even trump selection pressures in determining the dynamics of cultural evolution: a highly counter-intuitive hypothesis will fail to dominate a population, even if there are strong advantages to adopting it. These results suggest that one of the consequences of cultural transmission will be the adaptation of concepts and languages to match human inductive biases.

This work was supported by grants 0704034 and 0544705 from the National Science Foundation (to T.L.G. and M.L.K., respectively) and by a Discovery Project grant from the Australian Research council to S.L. We thank four anonymous reviewers for their comments.

ENDNOTES

[1] We refer to these constraints as *inductive biases* by analogy to the machine learning literature, in which the inductive bias of a learning algorithm is the set of assumptions that lead the algorithm to select one hypothesis over another (Mitchell 1997). By considering human learning as one such algorithm, we use inductive biases to refer to all factors, such as prior knowledge or expectations, that make ideas easier to learn or remember, whether they are derived from innate constraints or from experience.

[2] Note that PA refers to the agent from the previous generation in this equation, as the data are the utterances produced by the previous learner. We assume that PA and LA are the same across all learners, which amount to the assumption that the prior distribution $P(h)$ is also shared.

[3] It is worth emphasizing that this analysis only justifies a *connection* between the prior and the consequences of knowledge transmission: it does not indicate where the inductive biases expressed in the prior distribution of hypotheses come from, and thus does not in itself provide justification for the claims about modular cognitive architectures or innate domain-specific constraints on linguistic or ontological knowledge, which are associated with these positions (for further discussion of this point, see Griffiths & Kalish 2007).

[4] While much recent work applying these models (e.g. Nowak *et al.* 2002) has focused on the effects of frequency-dependent selection, we restrict ourselves here to the case where fitness does not depend on the composition of the population. Exploring the consequences of Bayesian learning in the context of frequency-dependent selection is an exciting direction for future work.

APPENDIX A

To derive the threshold on s, we observe that $\mathrm{d}x/\mathrm{d}t$ is a negative quadratic function in x, and takes positive value when $x=0$ ($\mathrm{d}x/\mathrm{d}t=q_{12}$) and negative values when $x=1$ ($\mathrm{d}x/\mathrm{d}t=-q_{21}s$). It follows that $\mathrm{d}x/\mathrm{d}t=0$ at exactly one point in [0,1]. When $s=1$, this point is π. If $\pi<0.5$, then we can ask what value of s is required such that the crossing point is greater than 0.5. The derivative of $\mathrm{d}x/\mathrm{d}t$ with respect to s is $-x^2+(1-q_{21})x$, which is positive at 0.5 provided $q_{21}<0.5$. Solving for s such that $\mathrm{d}x/\mathrm{d}t=0$ when $x=0.5$ thus gives us a threshold above which the equilibrium value of x will be greater than 0.5. Substituting 0.5 for x into 9 and solving for s gives equation (5.4). When $q_{21}>0.5$, the derivative of $\mathrm{d}x/\mathrm{d}t$ with respect to s at 0.5 is negative. Consequently, increasing s can only decrease $\mathrm{d}x/\mathrm{d}t$ at this point. We know that $\mathrm{d}x/\mathrm{d}t$ at 0.5 is negative when $s=1$, so no $s>1$ can result in an equilibrium in which the probability of hypothesis 1 is 0.5 or greater.

REFERENCES

Anderson, J. R. 1990 *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.

Anderson, J. R. 1991 The adaptive nature of human categorization. *Psychol. Rev.* **98**, 409–429. (doi:10.1037/0033-295X.98.3.409)

Anderson, J. R. & Milson, R. 1989 Human memory: an adaptive perspective. *Psychol. Rev.* **96**, 703–719. (doi:10.1037/0033-295X.96.4.703)

Ashby, F. G. & Alfonso-Reese, L. A. 1995 Categorization as probability density estimation. *J. Math. Psychol.* **39**, 216–233. (doi:10.1006/jmps.1995.1021)

Atran, S. 2001 The trouble with memes: inferences versus imitation in cultural creation. *Hum. Nat.* **12**, 351–381. (doi:10.1007/s12110-001-1003-0)

Atran, S. 2002 *In gods we trust: the evolutionary landscape of religion*. Oxford, UK: Oxford University Press.

Bartlett, F. C. 1932 *Remembering: a study in experimental and social psychology*. Cambridge, UK: Cambridge University Press.

Bickerton, D. 1981 *Roots of language*. Ann Arbor, MI: Karoma.

Boyd, R. & Richerson, P. J. 1985 *Culture and the evolutionary process*. Chicago, IL: University of Chicago Press.

Boyer, P. 1994 *The naturalness of religious ideas: a cognitive theory of religion*. Berkeley, CA: University of California Press.

Boyer, P. 1998 Cognitive tracks of cultural inheritance: how evolved intuitive ontology governs cultural transmission. *Am. Anthropol.* **100**, 876–889. (doi:10.1525/aa.1998.100.4.876)

Brighton, H. 2002 Compositional syntax from cultural transmission. *Artif. Life* **8**, 25–54. (doi:10.1162/106454602753694756)

Briscoe, E. (ed.) 2002 *Linguistic evolution through language acquisition: formal and computational models*, Cambridge, UK: Cambridge University Press.

Brown, D. E. 1991 *Human universals*. New York, NY: McGraw-Hill.

Buchanan, T. W. 2007 Retrieval of emotional memories. *Psychol. Bull.* **133**, 761–779. (doi:10.1037/0033-2909.133.5.761)

Caldwell, C. & Millen, A. E. 2008 Studying cumulative cultural evolution in the laboratory. *Phil. Trans. R. Soc. B* **363**, 3529–3539. (doi:10.1098/rstb.2008.0133)

Cavalli-Sforza, L. L. & Feldman, M. W. 1981 *Cultural transmission and evolution*. Princeton, NJ: Princeton University Press.

Chomsky, N. 1965 *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Comrie, B. 1981 *Language universals and linguistic typology*. Chicago, IL: University of Chicago Press.

Deacon, T. W. 1997 *The symbolic species: the co-evolution of language and the brain*. New York, NY: Norton.

DeGraff, M. (ed.) 1999 *Language creation and language change: creolization, diachrony, and development*, Cambridge, MA: MIT Press.

Feldman, J. 2000 Minimization of Boolean complexity in human concept learning. *Nature* **407**, 630–633. (doi:10.1038/35036586)

Flynn, E. 2008 Investigating children as cultural magnets: do young children transmit redundant information along diffusion chains? *Phil. Trans. R. Soc. B* **363**, 3541–3551. (doi:10.1098/rstb.2008.0136)

Galantucci, B. 2005 An experimental study of the emergence of human communication systems. *Cogn. Sci.* **29**, 737–767. (doi:10.1207/s15516709cog0000_34)

Garrod, S., Fay, N., Lee, J., Oberlander, J. & Macleod, T. 2007 Foundations of representation: where might graphical symbol systems come from? *Cogn. Sci.* **31**, 961–988. (doi:10.1080/03640210701703659)

Geman, S. & Geman, D. 1984 Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **6**, 721–741.

Gopnik, A. & Meltzoff, A. N. 1997 *Words, thoughts, and theories*. Cambridge, MA: MIT Press.

Greenberg, J. (ed.) 1963 *Universals of language*, Cambridge, MA: MIT Press.

Griffiths, T. L. & Kalish, M. L. 2005 A Bayesian view of language evolution by iterated learning. In *Proc. Twenty-Seventh Annual Conf. of the Cognitive Science Society* (eds B. G. Bara, L. Barsalou & M. Bucciarelli), pp. 827–832. Mahwah, NJ: Erlbaum.

Griffiths, T. L. & Kalish, M. L. 2007 A Bayesian view of language evolution by iterated learning. *Cogn. Sci.* **31**, 441–480.

Griffiths, T. L. & Tenenbaum, J. B. 2005 Structure and strength in causal induction. *Cogn. Psychol.* **51**, 354–384. (doi:10.1016/j.cogpsych.2005.05.004)

Griffiths, T. L., Steyvers, M. & Tenenbaum, J. B. 2007 Topics in semantic association. *Psychol. Rev.* **114**, 211–244. (doi:10.1037/0033-295X.114.2.211)

Griffiths, T. L., Christian, B. R. & Kalish, M. L. 2008 Using category structures to test iterated learning as a method for identifying inductive biases. *Cogn. Sci.* **32**, 68–107. (doi:10.1080/03640210701801974)

Hawkins, J. (ed.) 1988 *Explaining language universals*. Oxford, UK: Blackwell.

Henrich, J. & Boyd, R. 2002 Culture and cognition: why cultural evolution does not require replication of representations. *Cult. Cogn.* **2**, 87–112. (doi:10.1163/156853702320281836)

Hofbauer, J. & Sigmund, K. 1998 *Evolutionary games and population dynamics*. Cambridge, UK: Cambridge University Press.

Hudson-Kam, C. L. & Newport, E. L. 2005 Regularizing unpredictable variation: the roles of adult and child learners in language formation and change. *Lang. Learn. Dev.* **1**, 151–195. (doi:10.1207/s15473341lld0102_3)

Kalish, M., Lewandowsky, S. & Kruschke, J. 2004 Population of linear experts: knowledge partitioning and function learning. *Psychol. Rev.* **111**, 1072–1099. (doi:10.1037/0033-295X.111.4.1072)

Kalish, M. L., Griffiths, T. L. & Lewandowsky, S. 2007 Iterated learning: intergenerational knowledge transmission reveals inductive biases. *Psychon. Bull. Rev.* **14**, 288–294.

Kirby, S. 1999 *Function, selection and innateness: the emergence of language universals*. Oxford, UK: Oxford University Press.

Kirby, S. 2001 Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. *IEEE J. Evol. Comput.* **5**, 102–110. (doi:10.1109/4235.918430)

Kirby, S., Dowman, M. & Griffiths, T. L. 2007 Innateness and culture in the evolution of language. *Proc. Natl Acad. Sci. USA* **104**, 5241–5245. (doi:10.1073/pnas.0608222104)

Komarova, N. L. & Nowak, M. A. 2003 Language dynamics in finite populations. *J. Theor. Biol.* **221**, 445–457. (doi:10.1006/jtbi.2003.3199)

Marr, D. 1982 *Vision*. San Francisco, CA: W. H. Freeman.

Mesoudi, A. 2007 Using the methods of experimental social psychology to study cultural evolution. *J. Soc. Evol. Cult. Psychol.* **1**, 35–58.

Mesoudi, A. & Whiten, A. 2008 The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Phil. Trans. R. Soc. B* **363**, 3489–3501. (doi:10.1098/rstb.2008.0129)

Mitchell, T. M. 1997 *Machine learning*. New York, NY: McGraw Hill.

Nichols, S. 2004 A fragment of the genealogy of norms. *Sentimental Rules* **1**, 118–141. (doi:10.1093/0195169344.003.0006)

Norris, J. R. 1997 *Markov chains*. Cambridge, UK: Cambridge University Press.

Nosofsky, R. M., Gluck, M., Palmeri, T. J., McKinley, S. C. & Glauthier, P. 1994 Comparing models of rule-based classification learning: a replication and extension of Shepard, Hovland, and Jenkins (1961). *Mem. Cognit.* **22**, 352–369.

Nowak, M. A., Komarova, N. L. & Niyogi, P. 2001 Evolution of universal grammar. *Science* **291**, 114–118. (doi:10.1126/science.291.5501.114)

Nowak, M. A., Komarova, N. L. & Niyogi, P. 2002 Computational and evolutionary aspects of language. *Nature* **417**, 611–617. (doi:10.1038/nature00771)

Pinker, S. 2002 *The blank slate: the modern denial of human nature*. New York, NY: Viking.

Senghas, A., Kita, S. & Özyürek, A. 2004 Children creating core properties of language: evidence from an emerging sign language in Nicaragua. *Science* **305**, 1779–1782. (doi:10.1126/science.1100199)

Shepard, R. N., Hovland, C. I. & Jenkins, H. M. 1961 Learning and memorization of classifications. *Psychol. Monogr.* **75**, 1–42.

Shiffrin, R. M. & Steyvers, M. 1997 A model for recognition memory: REM: retrieving effectively from memory. *Psychon. Bull. Rev.* **4**, 145–166.

Smith, K. & Kirby, S. 2008 Cultural evolution: implications for understanding the human language faculty and its evolution. *Phil. Trans. R. Soc. B* **363**, 3591–3603. (doi:10.1098/rstb.2008.0145)

Smith, K., Kirby, S. & Brighton, H. 2003 Iterated learning: a framework for the emergence of language. *Artif. Life* **9**, 371–386. (doi:10.1162/106454603322694825)

Spelke, E. S., Breinlinger, K., Macomber, J. & Jacobson, K. 1992 Origins of knowledge. *Psychol. Rev.* **99**, 605–632. (doi:10.1037/0033-295X.99.4.605)

Sperber, D. 1985 Anthropology and psychology: towards an epidemiology of representations. *Man* **20**, 73–89. (doi:10.2307/2802222)

Sperber, D. 1996 *Explaining culture: a naturalistic approach*. Oxford, UK: Blackwell.

Sperber, D. & Claidiére, N. 2006 Why modeling cultural evolution is still such a challenge. *Biol. Theory* **1**, 20–22. (doi:10.1162/biot.2006.1.1.20)

Steels, L. 2003 Evolving grounded communication for robots. *Trends Cogn. Sci.* **7**, 308–312. (doi:10.1016/S1364-6613(03)00129-3)

Tenenbaum, J. B. & Griffiths, T. L. 2001 Generalization, similarity, and Bayesian inference. *Behav. Brain Sci.* **24**, 629–641. (doi:10.1017/S0140525X01000061)

Whiten, A. & Mesoudi, A. 2008 Establishing an experimental science of culture: animal social diffusion experiments. *Phil. Trans. R. Soc. B* **363**, 3477–3488. (doi:10.1098/rstb.2008.0134)