PROCEEDINGS OF THE ROYAL SOCIETY B

Proc. R. Soc. B (2011) **278**, 1773–1776
doi:10.1098/rspb.2011.0349
Published online 23 March 2011

Invited reply

Modelling minds as well as populations

Bentley *et al.* [1] present a spirited defence of the use of neutral models at the population level, and a demonstration that the three linguistic phenomena considered in our original article [2] can be captured through a different modification to the Wright–Fisher model than the one we considered. Their demonstration helps to reinforce our argument that these phenomena can be explained without the need to appeal to selection, showing that another simple neutral model can produce these effects. We see the key issues raised by their commentary as being whether there is any value in the novel connection that we identified between the Wright–Fisher model and cultural transmission by Bayesian agents, and whether this interpretation 'runs the risk of obscuring the advances made both through careful modifications and wider applications of this powerful model' (p. 1). We address these issues in turn.

To recapitulate our basic result, we showed that transmission of a probability distribution over a discrete set of alternatives by a sequence of Bayesian agents could be mathematically equivalent to the Wright–Fisher model [3,4], a classic model used in population genetics. Formally, each learner receives n observations of a set of variants (such as words or linguistic constructions), forms an estimate of the probability of each variant, summarized in a vector θ, and then generates n observations by sampling from this distribution. We showed that when the learners apply Bayesian inference with a particular prior distribution and choose the value of θ with highest posterior probability, the frequencies of the variants follow the dynamics of the Wright–Fisher model. We then used the connection to Bayesian inference to introduce a more flexible variant of this model, which we applied to the linguistic phenomena mentioned above.

We agree with Bentley *et al.* [1] that the Bayesian interpretation of this model is more complex than the original model, in which cultural transmission is described simply in terms of copying variants with some chance of 'mutation', by direct analogy to biological transmission. This complexity results from considering the cognitive mechanisms that underlie each cultural transmission event, and treating them as being more sophisticated than error-prone copying. Our interest in these mechanisms is partly a consequence of our bias as cognitive scientists, but also reflects our expectation that the mechanisms of cultural transmission are going to be different from (and potentially far more complex than) the mechanisms of biological transmission. In the case of the Wright–Fisher model, we believe that our mathematical analysis has several implications that make this extra complexity worthwhile.

As a general methodological point, we first note that mathematical results connecting different models are valuable not just because they provide new interpretations of those models, but because they extend the tools that are available for analysing them. In this case, we provide a link to a broader class of 'iterated learning' models that have been used to model language evolution [5,6]. In these models, a sequence of agents each hear a set of utterances, form a hypothesis about the language and then generate the utterances that are heard by the next agent. When the agents use Bayesian inference, the outcome of this process is well understood. In particular, if agents choose hypotheses by sampling from their posterior distribution, over time the probability an agent selects a particular hypothesis converges to the prior probability of that hypothesis [7]. This process can be shown to be a form of Gibbs sampling, a Markov chain Monte Carlo algorithm that is widely used in Bayesian statistics [7]. These results can potentially provide new insight into the Wright–Fisher model: standard asymptotic analyses of this model use diffusion approximations (see [8]), but establishing the link to iterated learning (and Gibbs sampling) indicates that there is a closely related class of models where the asymptotic behaviour is exactly known. If agents select the hypothesis with highest posterior probability, as is required to establish equivalence to the Wright–Fisher model, then iterated learning becomes equivalent to a different statistical inference algorithm, known as the stochastic expectation-maximization algorithm [7]. Again, asymptotic analyses exist for this algorithm (e.g. [9]), and showing that Wright–Fisher is an instance of this algorithm that has the potential to allow mathematical results to generalize in both directions.

Beyond these mathematical implications, providing a link to Bayesian inference establishes a connection between existing work using the Wright–Fisher model in cultural evolution and a growing literature in cognitive science on Bayesian models of cognition. While Bentley *et al.* [1] emphasize the recent disenchantment with rational models of decision-making in behavioural economics, there has been a parallel growth of interest in rational models of cognition in cognitive psychology [10–12]. Bayesian inference provides a way to answer a key question that comes up in describing human learning and memory, indicating how the expectations of an agent combine with the observed data to yield a conclusion. The prior distribution that is used in Bayesian inference expresses those factors other than the data that influence the conclusions that agents reach, including innate dispositions, schemas established through past experience, and prior knowledge about a particular situation. In formal

The accompanying comment can be viewed at http://dx.doi.org/doi:10.1098/rspb.2010.2581.

Received 15 February 2011
Accepted 4 March 2011